Modeling Symbolic Music with Natural Language Processing Approaches

Analyzing and Experimenting with Multiple Levels of Representations

Dinh-Viet-Toan LE

PhD Thesis Defense

November 3, 2025

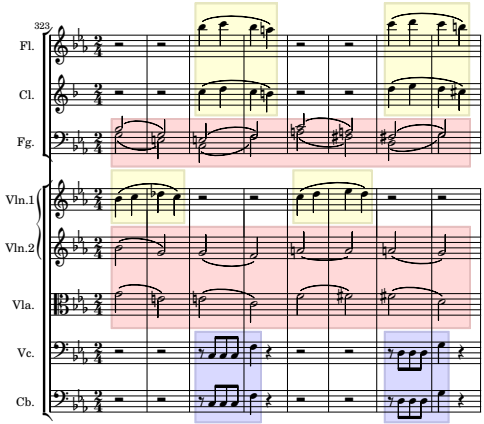
Xavier HINAUT	Chargé de recherche Inria, Inria Bordeaux	Reviewer
Cheng-Zhi Anna HUANG	Associate Professor, MIT	Reviewer
Emmanouil BENETOS	Reader & Director of research, QMUL	Examiner
Chloé BRAUD	Chargée de recherche CNRS, IRIT	Examiner
Marius BILASCO	Professeur des universités, CRIStAL	Examiner
Patrick BAS	Directeur de recherche CNRS, CRIStAL	Jury president
Louis BIGO	Professeur des universités, LaBRI	Co-director
Marc TOMMASI	Professeur des universités, CRIStAL	Co-director
Mikaela KELLER	Maîtresse de conférence, CRIStAL	Co-supervisor

sound-check

Université de Lille

CRIStAL

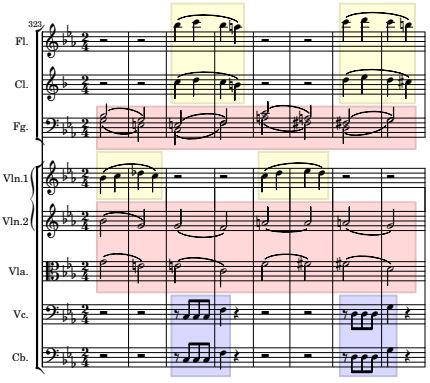
Music & language



L.v. Beethoven, *Symphony No. 5*, Mvnt. 1. Bars 323–330.

Music & language

"A **musical phrase** is passed through the instruments like a **dialog**."



L.v. Beethoven, *Symphony No. 5*, Mvnt. 1. Bars 323–330.

- "Call-and-response"
- "Musical discourse"
- "A composer's vocabulary"
- ...
- **"Musical *language*"**

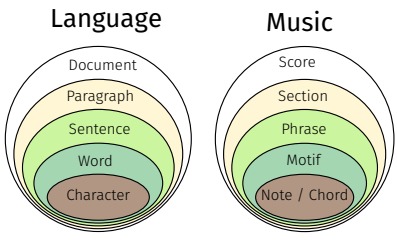
Jackendoff, *Parallels and nonparallels between language and music, 2009.*

Music & language: some fundamental similarities...

Modalities

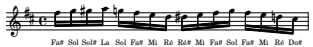
	Auditory	Written
Language	Speech	Text
Music	Audio	Score

Segmentation



Sequential representation

```
<Model_><_ing> <music_><_al> <content> <with> <NLP>
```



"Grammar"

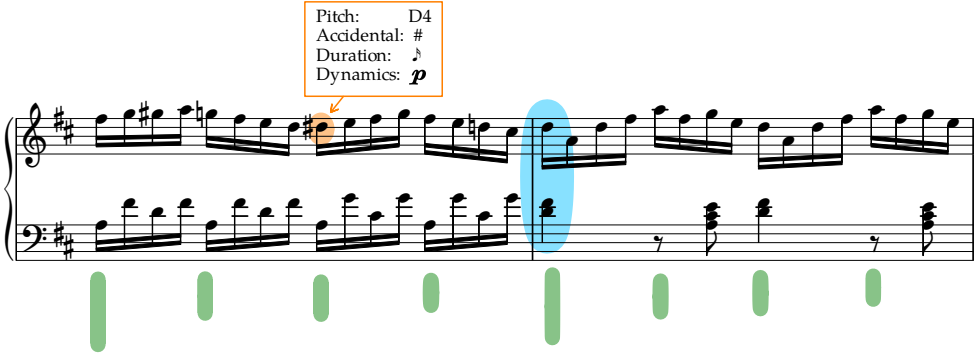
- Grammatical trees
- Tonal theory (GTTM [Lerdahl & Jackendoff 1987])

Expectancy

- `The cat is eating the [X]`
- Functional harmony no-cadence tonic

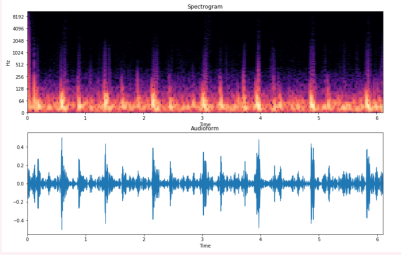
Music & language: ...and major differences

- Temporal dimension in language & music
- Polyphony in music
- Polymorphism of musical symbols



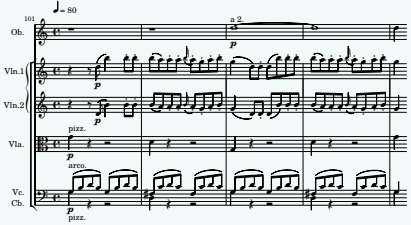
Symbolic music

Audio



- Waveforms, spectrograms...
- Timbre, performance expressivity...

Symbolic music



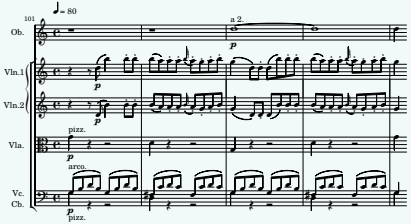
- Score (sheet music):
  - Notes, chords, instruments, dynamics, articulations...

Symbolic music

Audio

- Waveforms, spectrograms...
- Timbre, performance expressivity...

Symbolic music



- Score (sheet music):
  - Notes, chords, instruments, dynamics, articulations...

Natural Language Processing and *Music Information Retrieval*

Natural Language Processing (NLP)

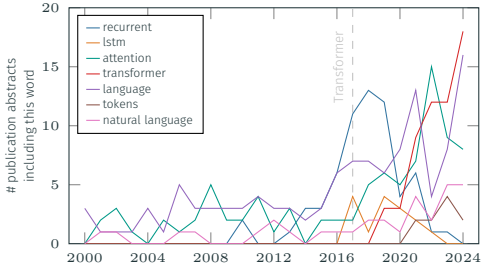
- Computer science applied to linguistics
- Interaction between computer and *written* human natural language

Music Information Retrieval (MIR)

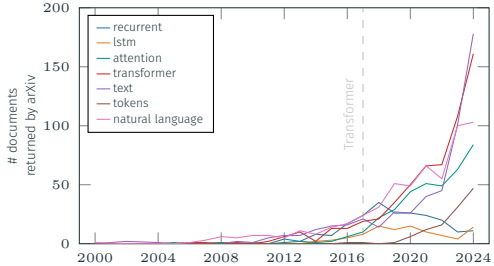
- Computer science applied to music
- Interaction between computer and (symbolic) music

NLP tools in symbolic MIR research

Number of ISMIR papers which include NLP-related terms in their abstract

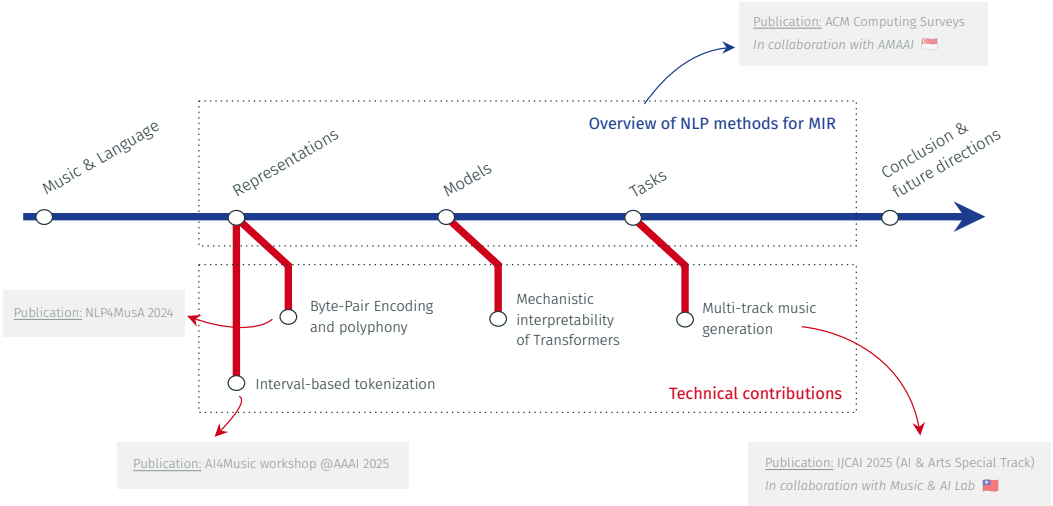


Number of MIR arXiv papers involving NLP-related terms

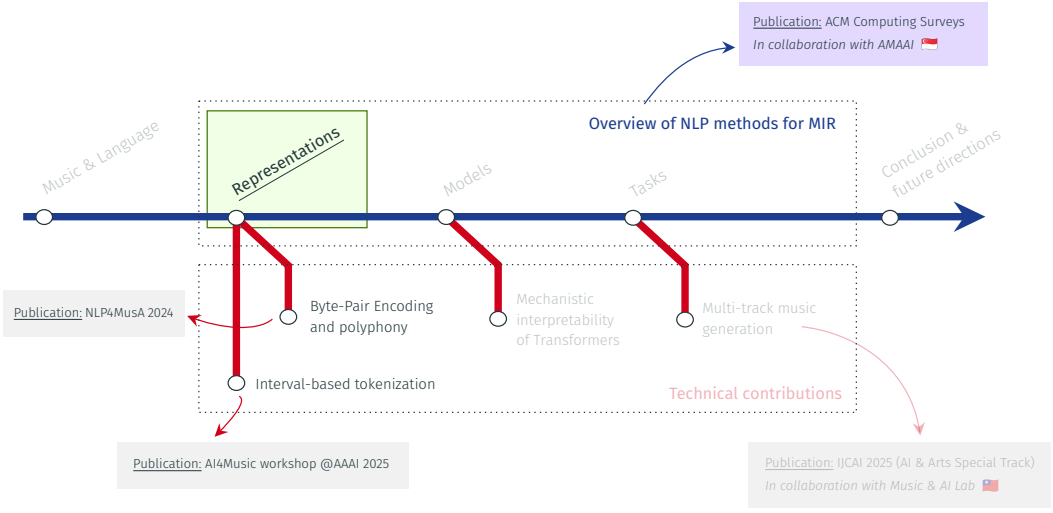


Le & al., *Natural Language Processing Methods for Symbolic Music Generation and Information Retrieval: A Survey*, ACM Computing Surveys 2025. **In collaboration with AMAAI (Singapore University of Technology and Design).**

Outline


Publication: ACM Computing Surveys
In collaboration with AMAAI
Overview of NLP methods for MIR
Music & Language
Representations
Models
Tasks
Conclusion & future directions
Publication: NLP4MusA 2024
Byte-Pair Encoding and polyphony
Mechanistic interpretability of Transformers
Multi-track music generation
Interval-based tokenization
Technical contributions
Publication: AI4Music workshop @AAAI 2025
Publication: IJCAI 2025 (AI & Arts Special Track)
In collaboration with Music & AI Lab

Outline – Sequential representations



Tokenization in text

Tokenization

Representing a complex content (*e.g.* text, music) into a **sequence** of elements for computational processing.

Tokenization in text

Tokenization

Representing a complex content (*e.g.* text, music) into a **sequence** of elements for computational processing.

`I love singing but it is exhausting`

Word-level tokenization

→ `<I>, <love>, <singing>, <but>, <it>, <is>, <exhausting>`

Character-level tokenization

→ `<I>, <_>, <l>, <o>, <v>, <e>, <_>, <s>, <i>, <n>, ...`

Subword-level tokenization

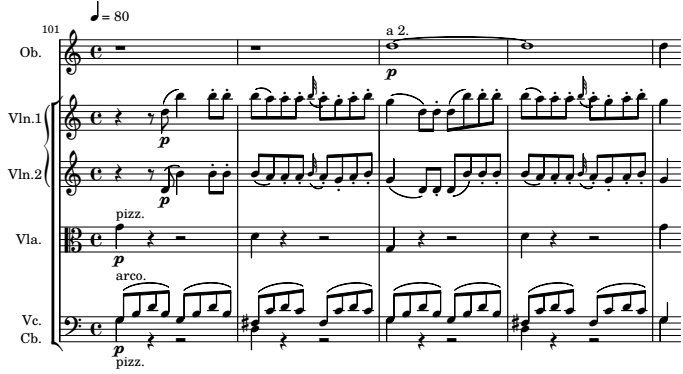
→ `<I>, <love>, <sing>, <_ing>, <but>, <it>, <is>, <exhaust>, <_ing>`

What does a musical score encode?

Tempo

Rhythm

Dynamics



Articulations

Pitch

Polyphony

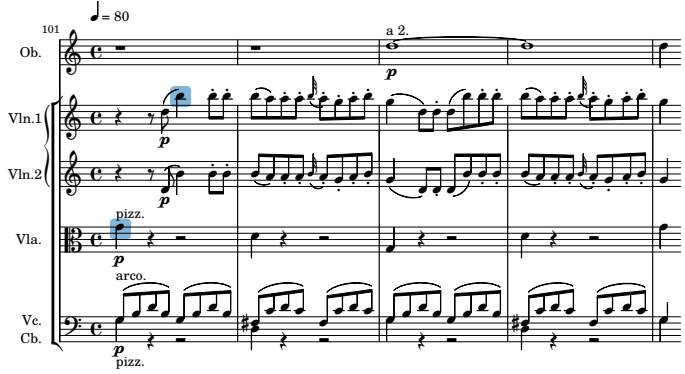
Structure

What does a musical score encode?

Tempo

Rhythm

Dynamics



Articulations

Pitch

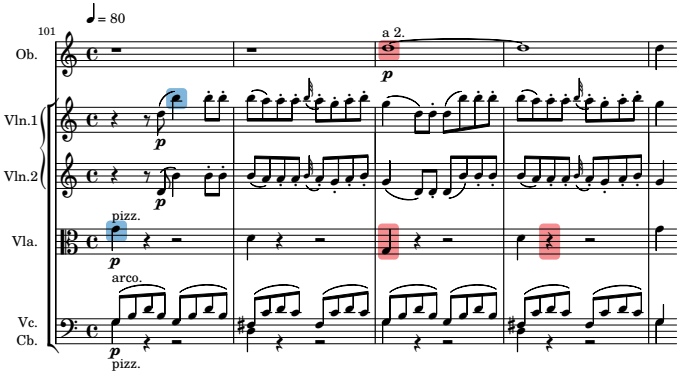
Polyphony

Structure

What does a musical score encode?

Tempo

Rhythm



Dynamics

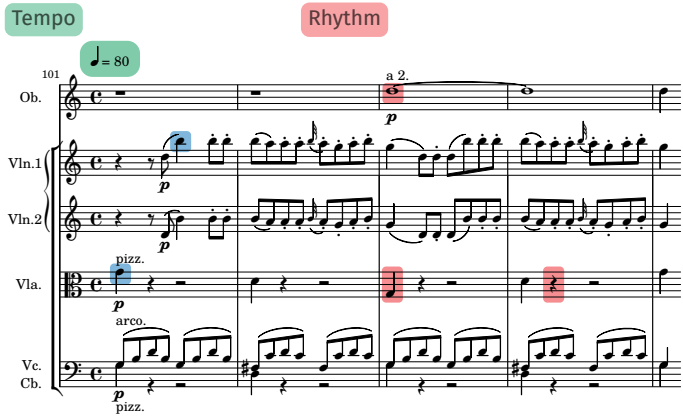
Articulations

Pitch

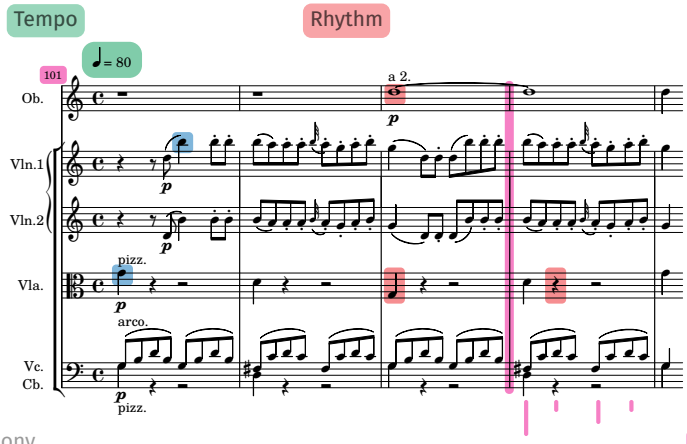
Polyphony

Structure

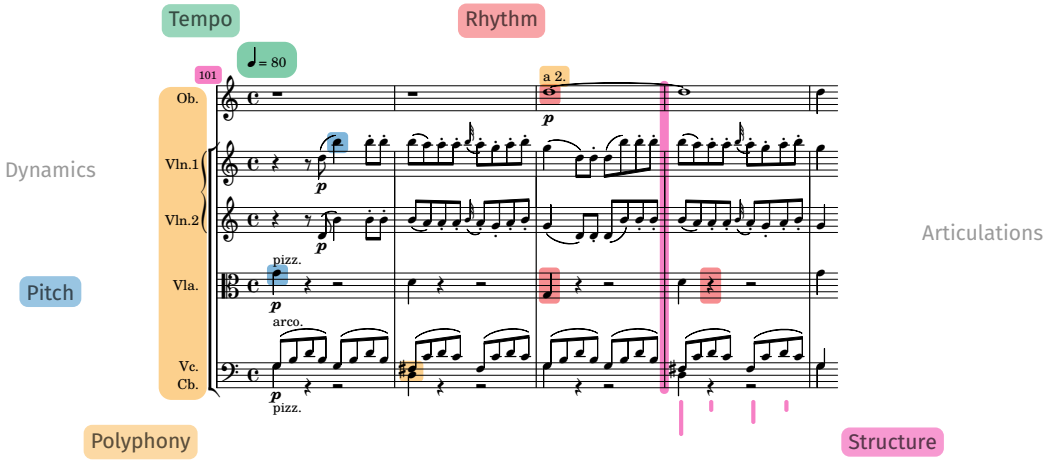
What does a musical score encode?



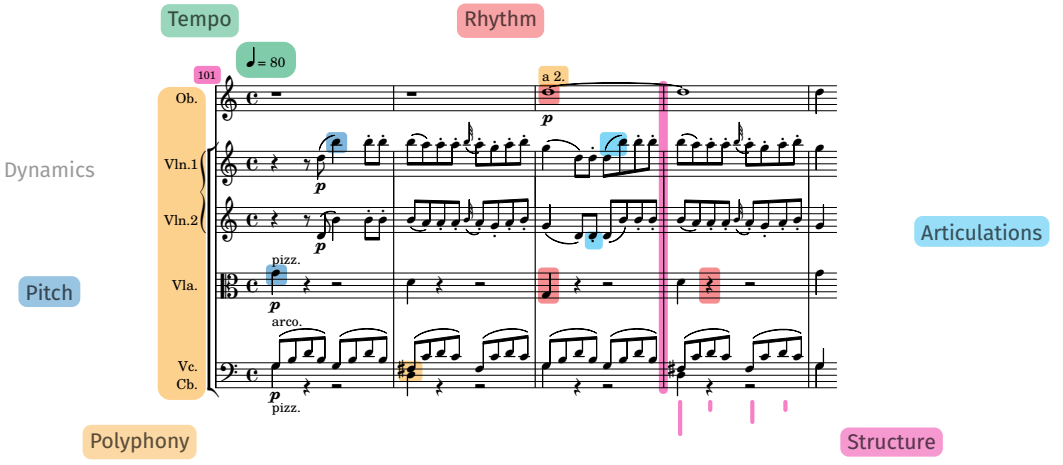
What does a musical score encode?



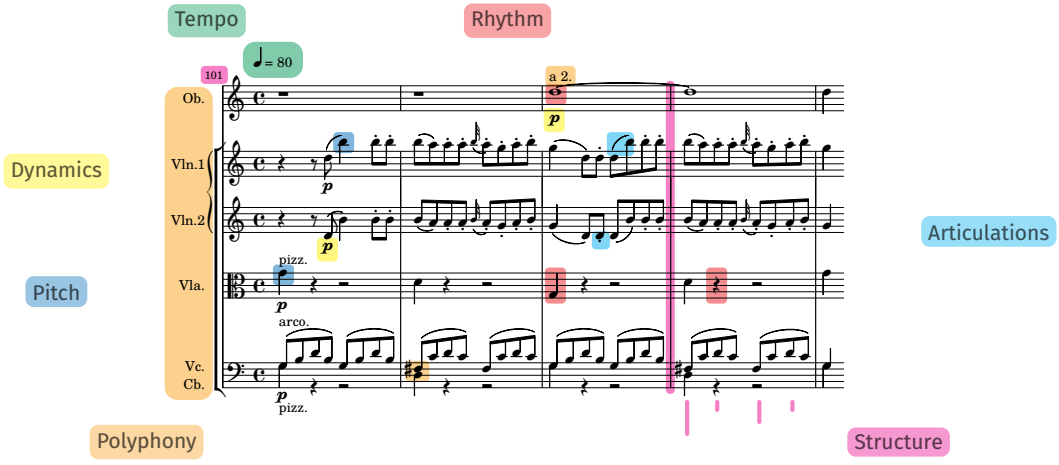
What does a musical score encode?



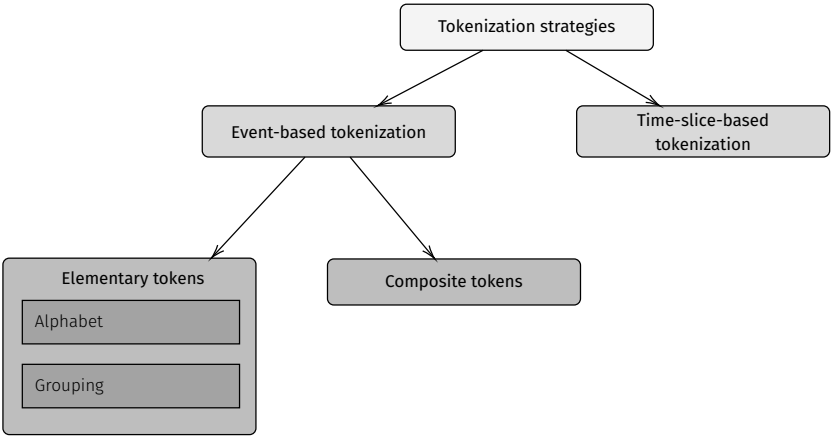
What does a musical score encode?



What does a musical score encode?

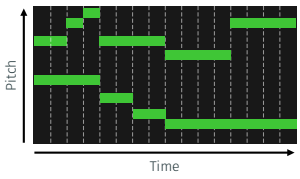


A taxonomy of tokenization strategies for symbolic music

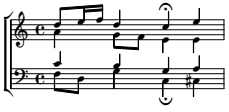


How to tokenize music?

Pianoroll slices

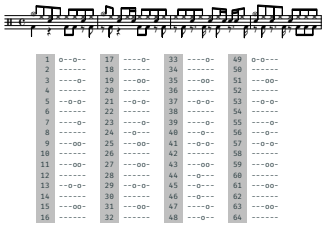


MIDI-VAE [Brunner & al. 2018]
Harmony Transformer [Chen & al. 2019]
RNBert [Sailor 2024]



```
D5, __, E5, F5, D5, __, __, __, C5, __, __, __, E5
A4, __, __, G4, __, F4, __, E4, __, __, __, __, E4
C4, __, __, B3, __, __, __, G3, __, __, __, __, A3
F3, __, D3, __, G3, __, __, __, C2, __, __, __, C#2
1,  2,  3,  4,  1,  2,  3,  4,  1,  2,  3,  4,  1
0,  0,  0,  0,  0,  0,  0,  0,  1,  1,  1,  1,  0
```

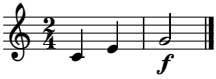
DeepBach [Hadjeres & al. 2017]



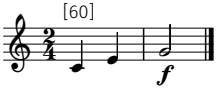
1	o---o--	17	----o-	33	----o-	49	o-o---
2	------	18	------	34	------	50	------
3	----o-	19	---oo-	35	---oo-	51	---oo-
4	------	20	------	36	------	52	------
5	--o-o-	21	--o-o-	37	--o-o-	53	--o-o-
6	------	22	------	38	------	54	------
7	----o-	23	----o-	39	----o-	55	----o-
8	------	24	--o---	40	---o--	56	---o--
9	---oo-	25	---oo-	41	--o-o-	57	--o-o-
10	------	26	------	42	------	58	------
11	---oo-	27	---oo-	43	---oo-	59	---oo-
12	------	28	------	44	--o---	60	------
13	--o-o-	29	--o-o-	45	--o---	61	---oo-
14	------	30	------	46	--o---	62	------
15	---oo-	31	---oo-	47	----o-	63	---oo-
16	------	32	------	48	---o--	64	------

Drums tokenization [Zhang & al. 2023]

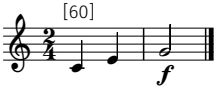
How to tokenize music?



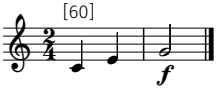
How to tokenize music?



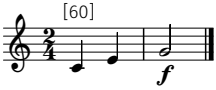
How to tokenize music?

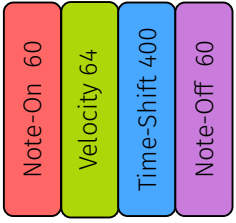


How to tokenize music?

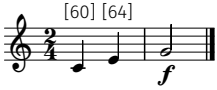


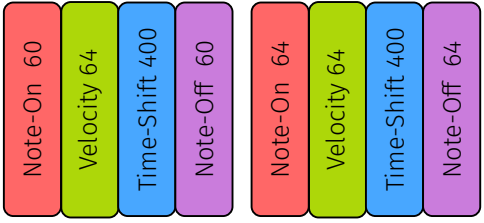
How to tokenize music?



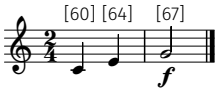


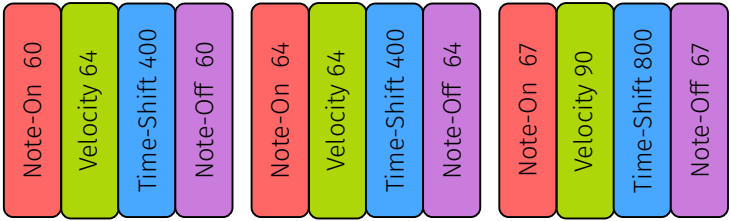
How to tokenize music?





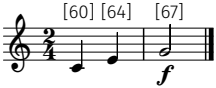
How to tokenize music?

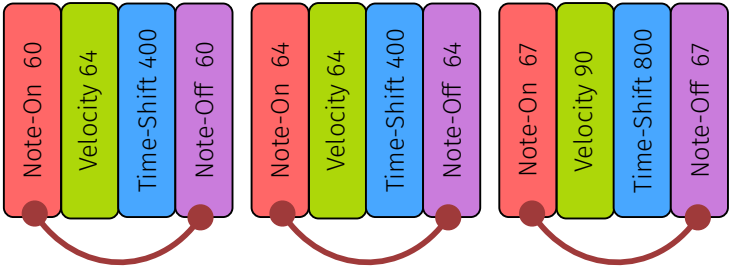




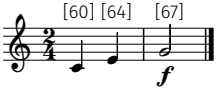
MIDI-like: Huang & al., *Music Transformer: Generating Music with Long-term Structure*, ICLR 2019

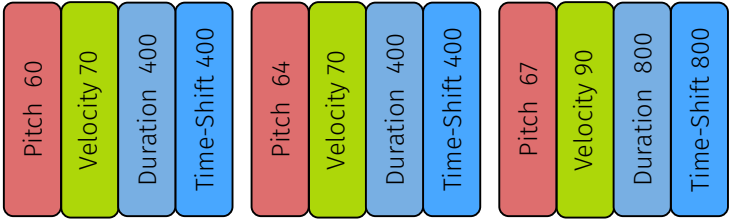
How to tokenize music?





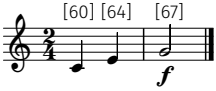
How to tokenize music?

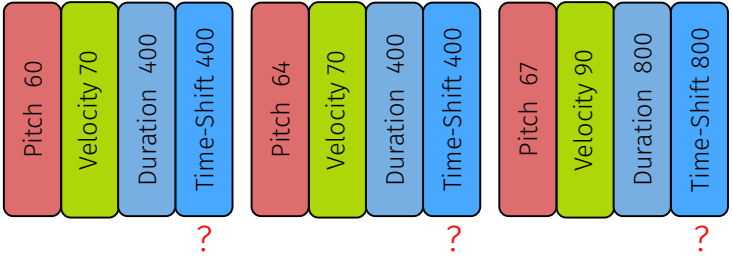




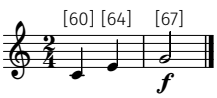
Structured: Hadjares & al., *The Piano Inpainting Application*, 2021

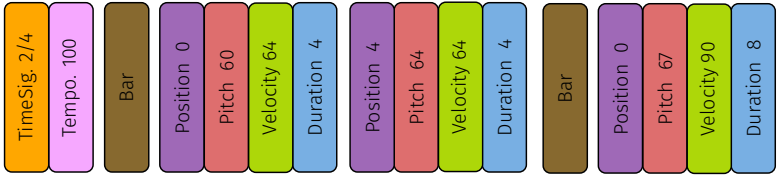
How to tokenize music?





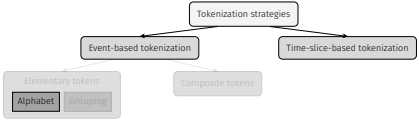
How to tokenize music?

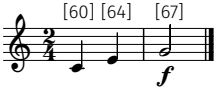


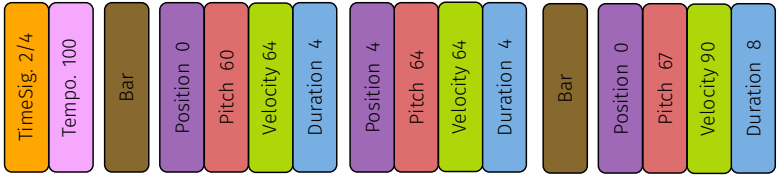


REMI: Huang & al., *Pop Music Transformer*, MM 2020

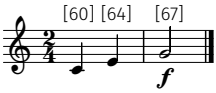
How to tokenize music?

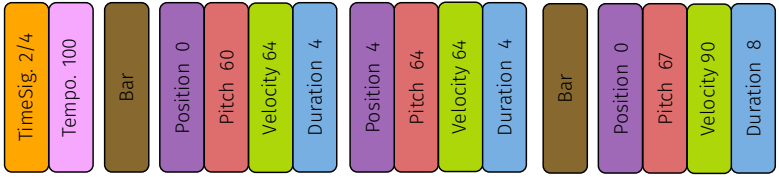






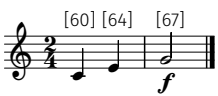
How to tokenize music?

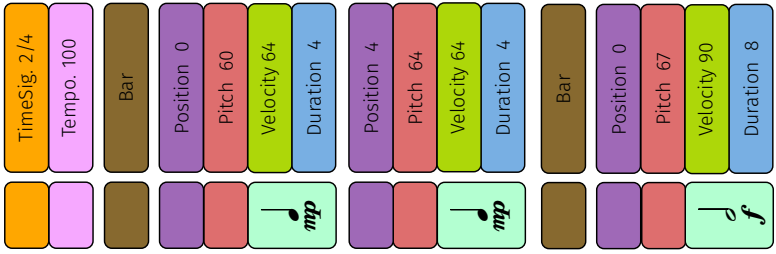




Grouping tokens together?

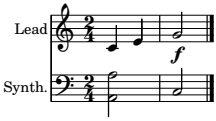
How to tokenize music?



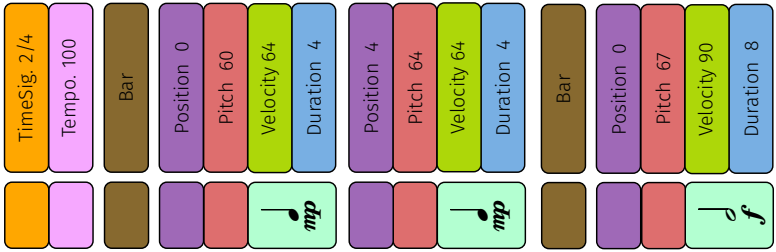


Rule-based ; **n-gram**: Conklin & al., 1995; **Byte-Pair Encoding**: Fradet & al., 2023

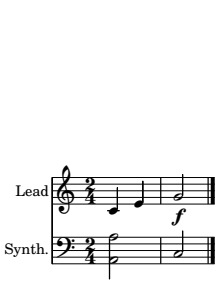
How to tokenize music?

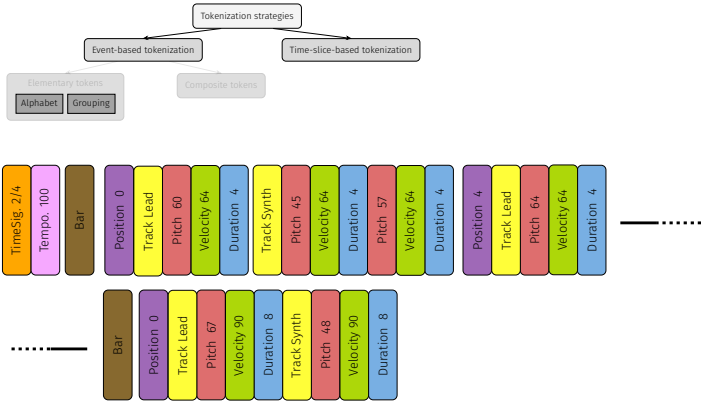


Multi-track?



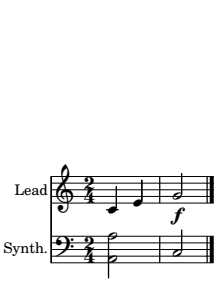
How to tokenize music?

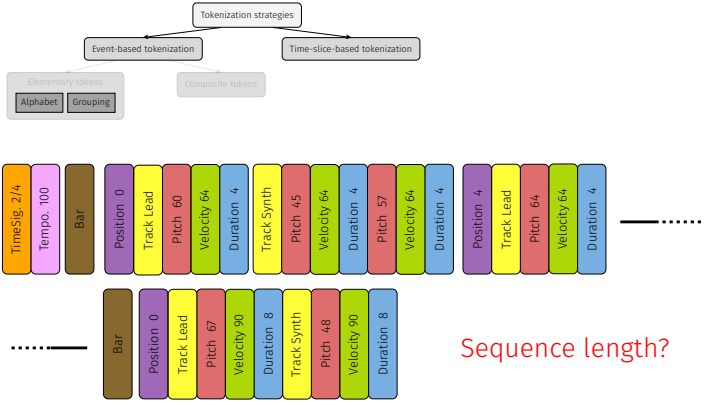




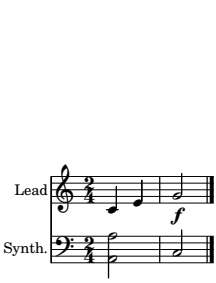
REMI+: von Rutte & al., *FIGARO: Generating Symbolic Music with Fine-Grained Artistic Control*, ICLR 2023

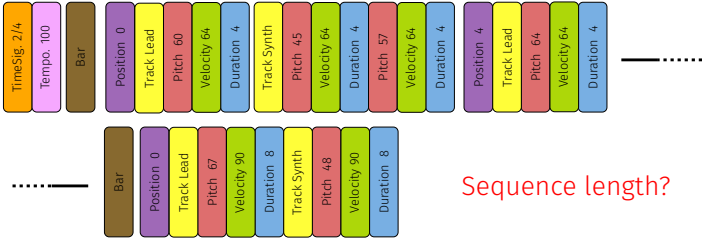
How to tokenize music?





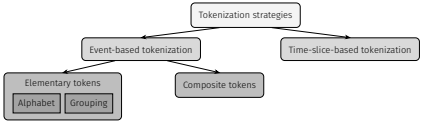
How to tokenize music?

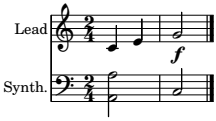


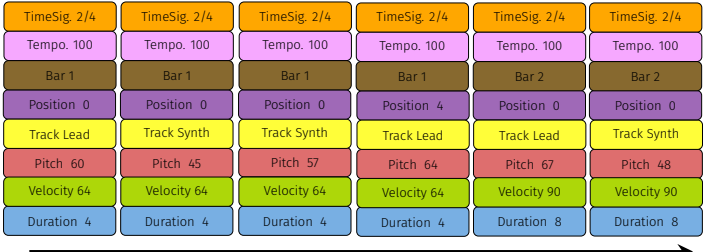


Sequence length?

How to tokenize music?

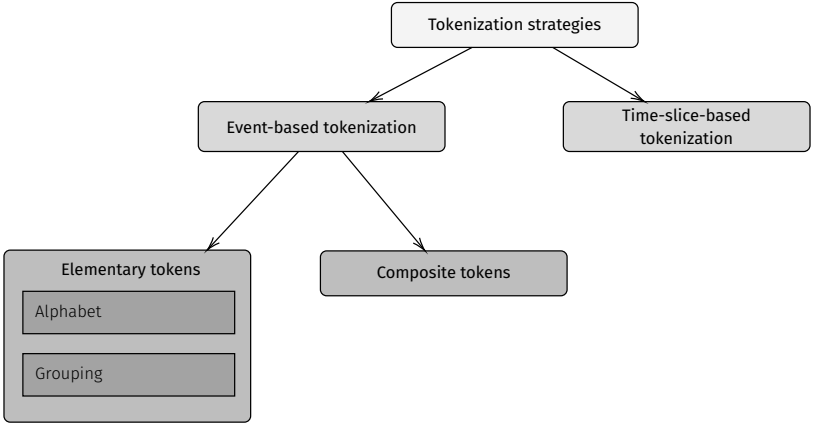






Octuple: Zeng & al., *MusicBERT: Symbolic Music Understanding with Large-Scale Pre-Training*, ACL 2021

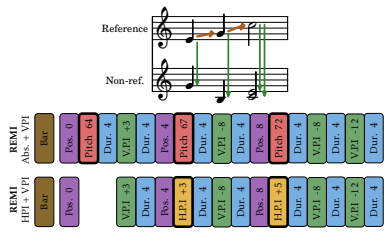
A taxonomy of tokenization strategies for symbolic music



Le & al., *Natural Language Processing Methods for Symbolic Music Generation and Information Retrieval: A Survey*, ACM Computing Surveys 2025. **In collaboration with AMAAI (Singapore University of Technology and Design).**

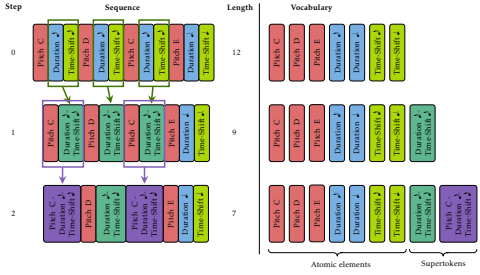
Technical contributions – Tokenization expressiveness

Alphabet: Interval-based tokenization



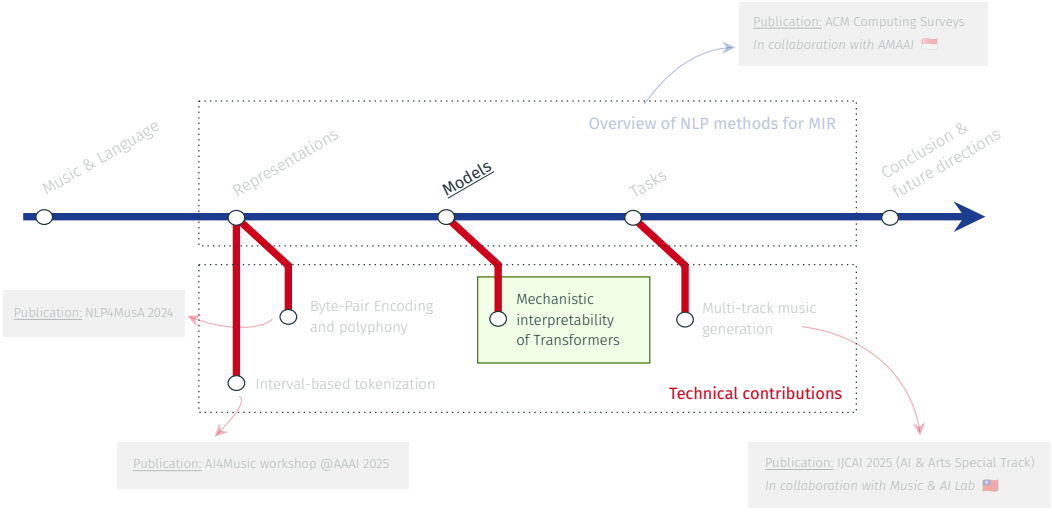
Le & al., *Evaluating Interval-based Tokenization for Pitch Representation in Symbolic Music Analysis*, AI4Music workshop at AAAI 2025

Grouping: Byte-pair encoding & polyphony

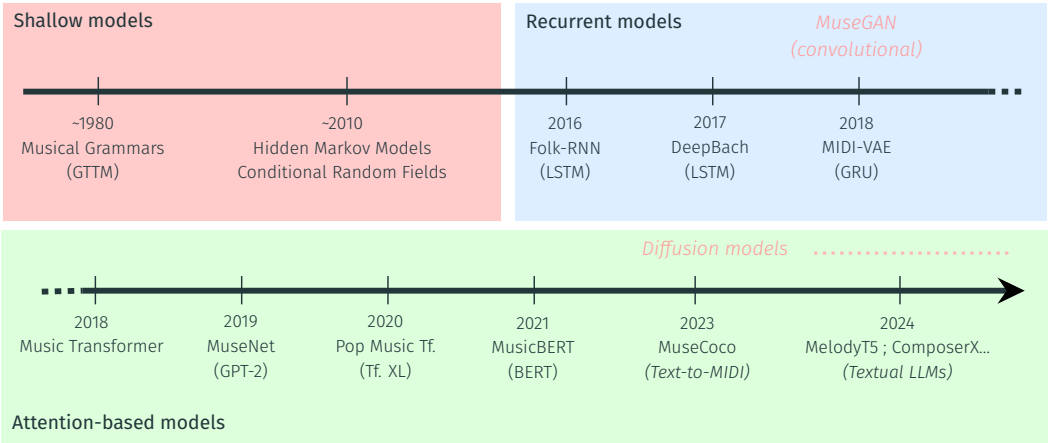


Le & al., *Analyzing Byte-Pair Encoding on Monophonic and Polyphonic Symbolic Music: A Focus on Musical Phrase Segmentation*, NLP4MusA 2024

Outline – Models

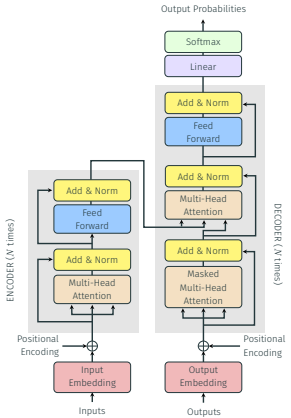


NLP models for symbolic music processing



Attention mechanism

- Build new representations of an input token sequence
- Weight the importance of each token according to its relevance to the others



Vaswani & al., *Attention is All You Need*, 2017

Model interpretability in MIR

Model interpretability *(or explainability)*

- Make a model's decision process transparent, faithful, and interpretable
- *Human-centered* explanations

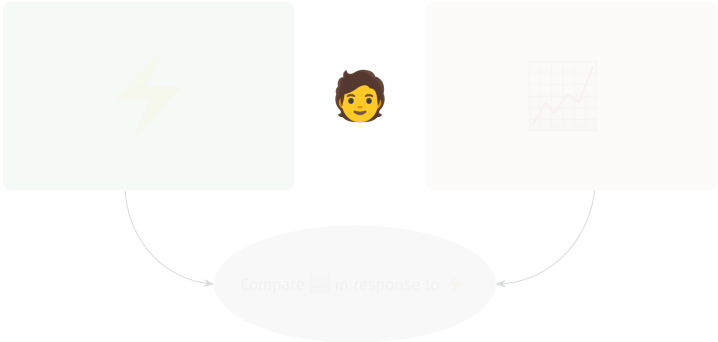
In Music Information Retrieval:

- **Chain-of-thoughts** for ABC Notation-based models [Zhou & al. 2024]
- Attention mechanism **visualization** in music generation [Huang & al. 2018]

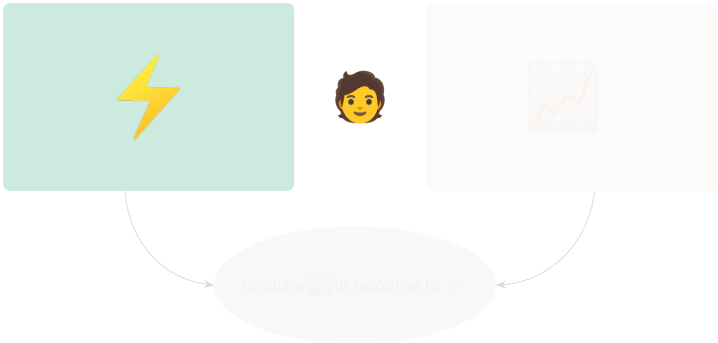
⇒ Our approach: **Analyzing** the attention mechanism in the context of functional harmonic analysis

A methodology to interpret attention with regards to musical characteristics

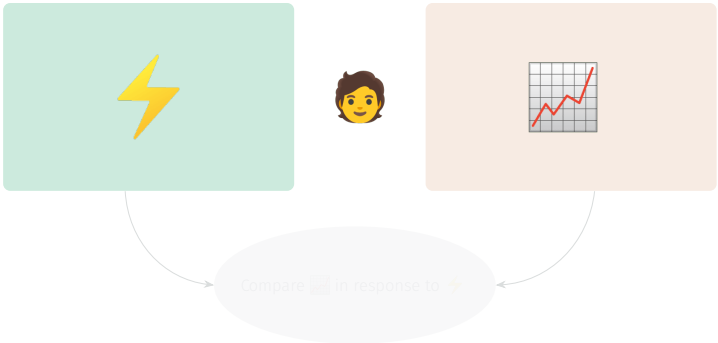
A methodology to interpret attention with regards to musical characteristics



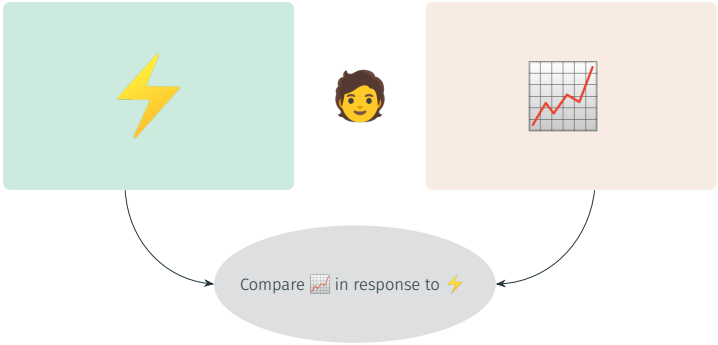
A methodology to interpret attention with regards to musical characteristics



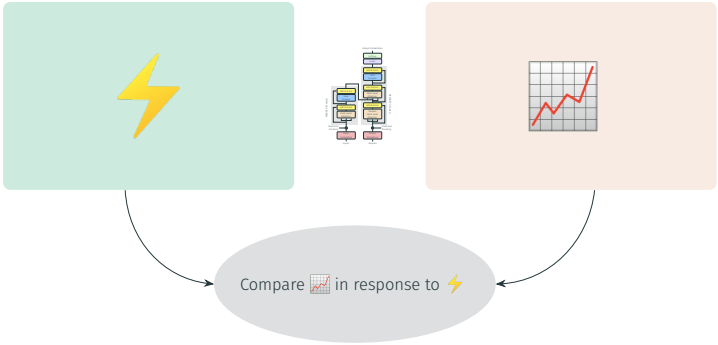
A methodology to interpret attention with regards to musical characteristics



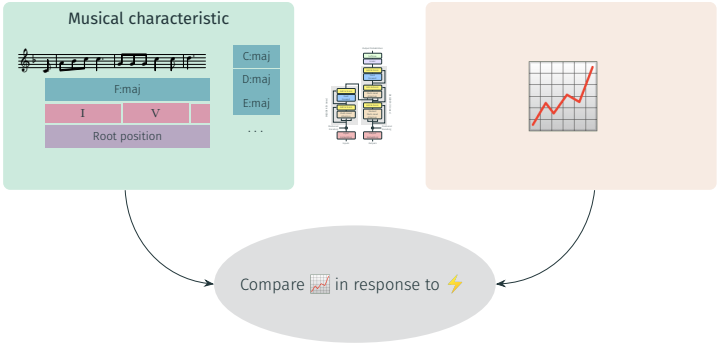
A methodology to interpret attention with regards to musical characteristics



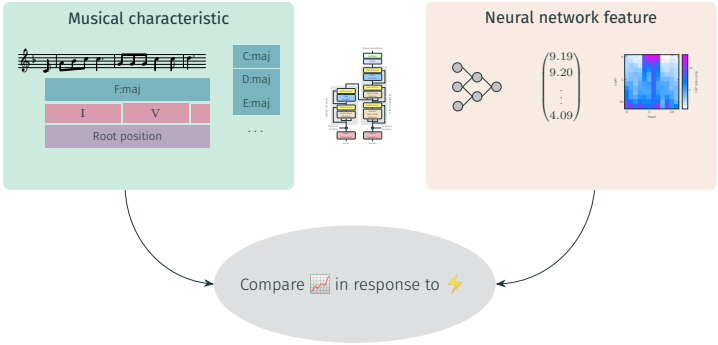
A methodology to interpret attention with regards to musical characteristics



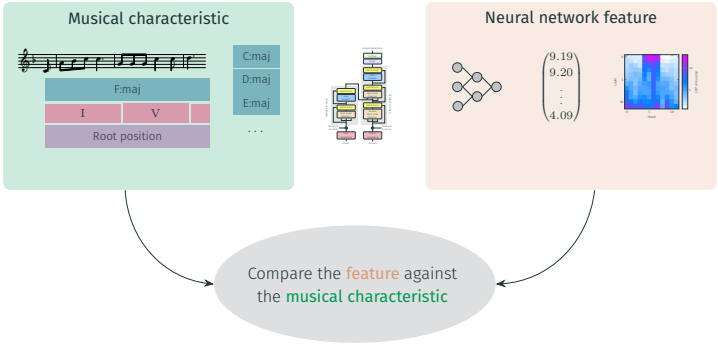
A methodology to interpret attention with regards to musical characteristics



A methodology to interpret attention with regards to musical characteristics

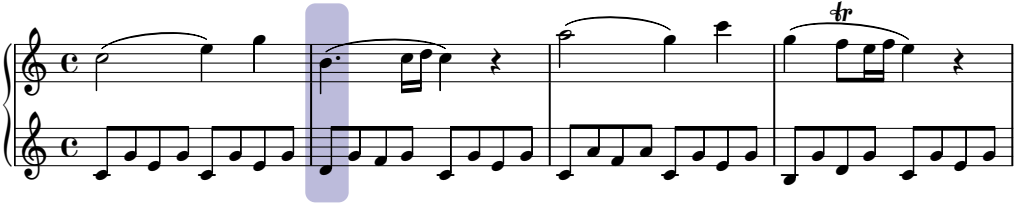


A methodology to interpret attention with regards to musical characteristics



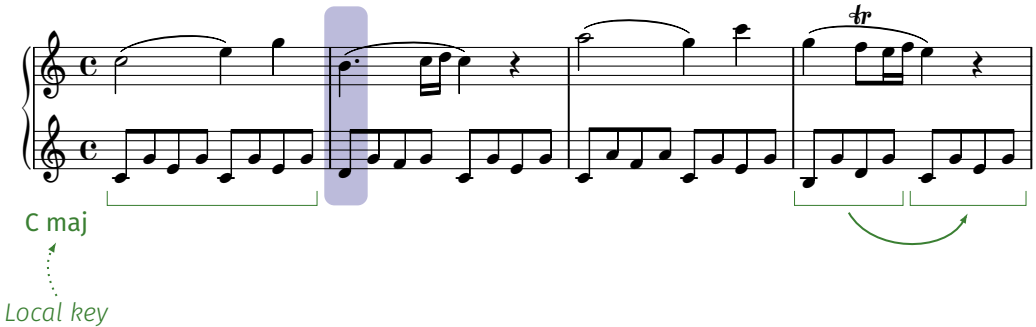
Functional Harmonic Analysis (*or Roman Numeral Analysis*)

Is there a human attention mechanism somewhere? mozartk545-lang-lang



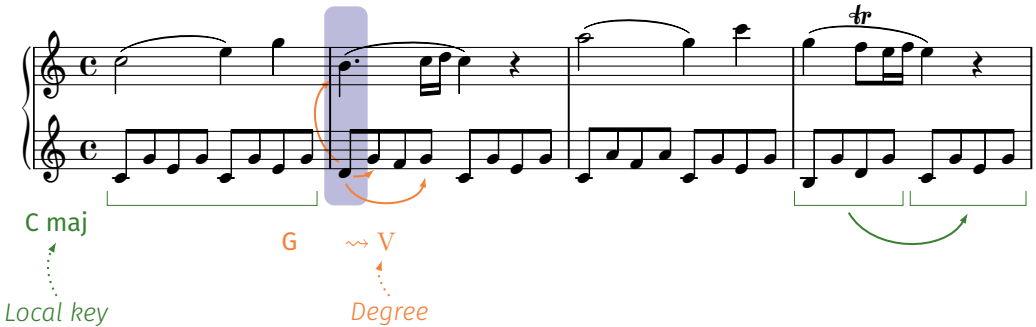
Functional Harmonic Analysis (*or Roman Numeral Analysis*)

Is there a human attention mechanism somewhere? mozartk545-lang-lang



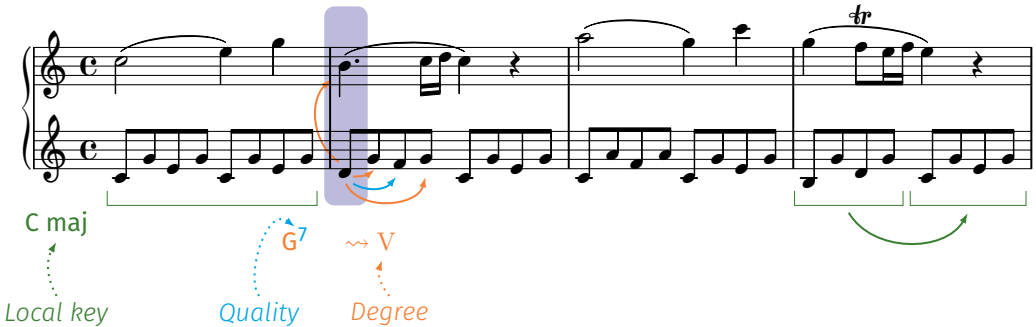
Functional Harmonic Analysis (*or Roman Numeral Analysis*)

Is there a human attention mechanism somewhere? mozartk545-lang-lang



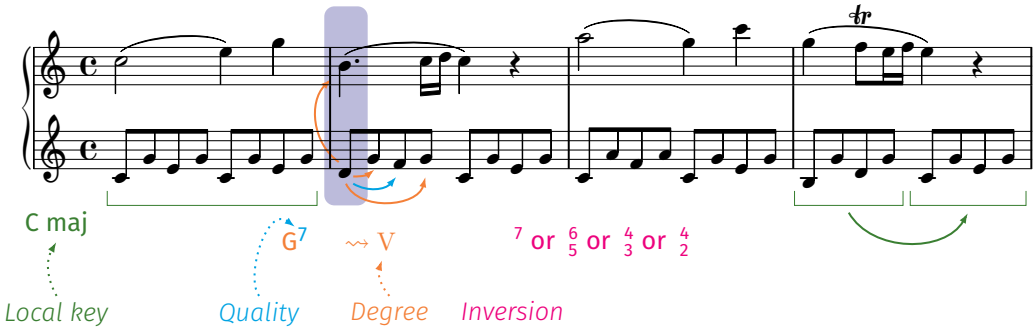
Functional Harmonic Analysis (*or Roman Numeral Analysis*)

Is there a human attention mechanism somewhere? mozartk545-lang-lang



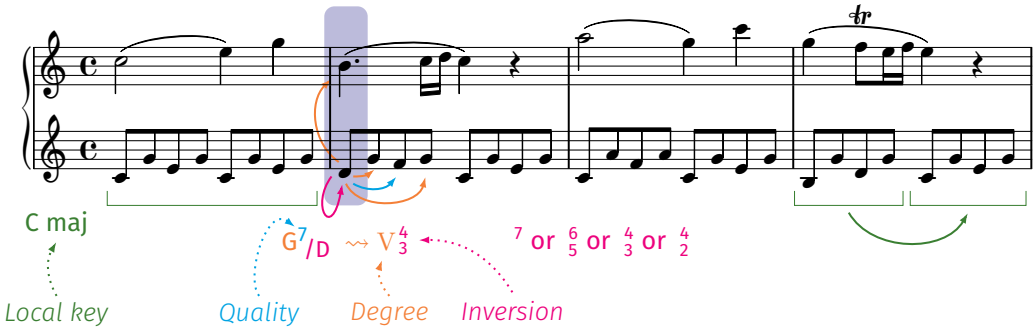
Functional Harmonic Analysis (*or Roman Numeral Analysis*)

Is there a human attention mechanism somewhere? mozartk545-lang-lang



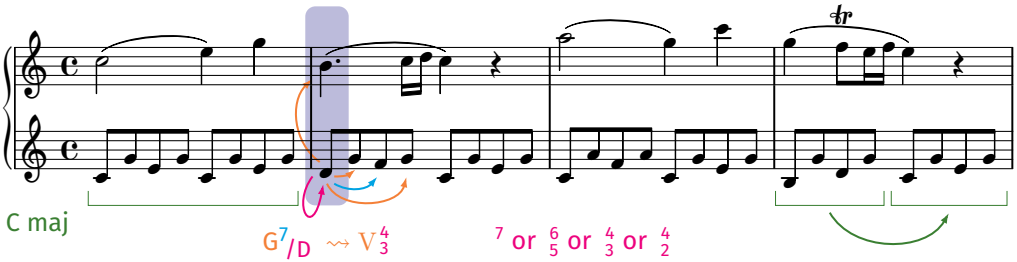
Functional Harmonic Analysis (*or Roman Numeral Analysis*)

Is there a human attention mechanism somewhere? mozartk545-lang-lang



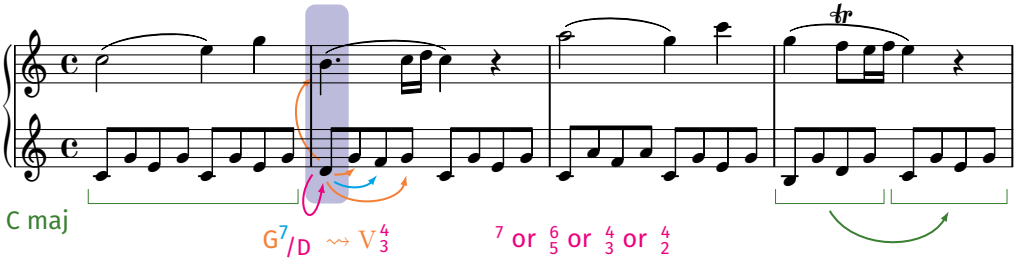
Functional Harmonic Analysis (*or Roman Numeral Analysis*)

Is there a human attention mechanism somewhere? mozartk545-lang-lang

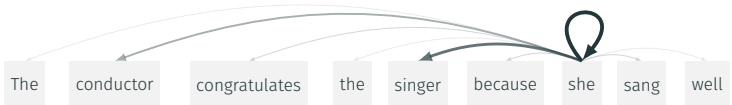


Functional Harmonic Analysis (*or Roman Numeral Analysis*)

Is there a human attention mechanism somewhere? mozartk545-lang-lang



≈ “attention arrows” in text?

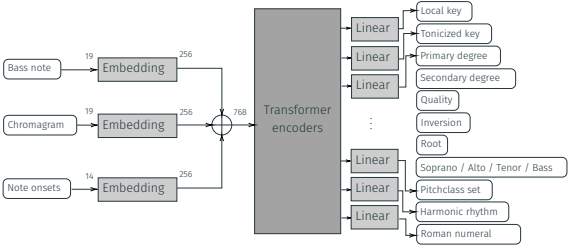


Automatic Functional Harmonic Analysis

Representation

- Time-slice-based tokenization: quantization at the 

Model architecture



Transformer encoders

→ Stacked layers of multi-head self-attention[1]

- Bi-directional
- 12 layers
- 12 **attention heads** per layer

[1] Vaswani & al., *Attention is All You Need*, 2017

Training

- End-to-end training
- 596 pieces, 28M tokens [Nápoles López & al., 2021]
- Era: baroque → early romantic
- Instruments: solo piano → string quartet

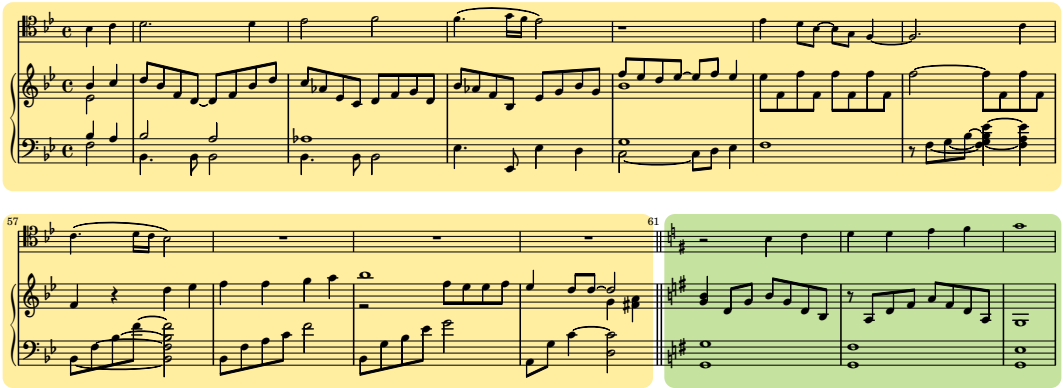
Performances (accuracy)

Model	Key	Degree	Quality	Inversion	Root	Global	# params
AugmentedNet v1.9.1	82.2	67.0	79.7	78.8	83.0	–	**105k**
RNBert	**82.5**	**85.9**	86.5	87.2	–	–	109M
Our model (mono-task)	79.5	81.7	90.0	89.7	92.0	87.5	–
Our model (multi-task)	81.7	84.3	**90.9**	**90.3**	**92.8**	**88.6**	94.9M

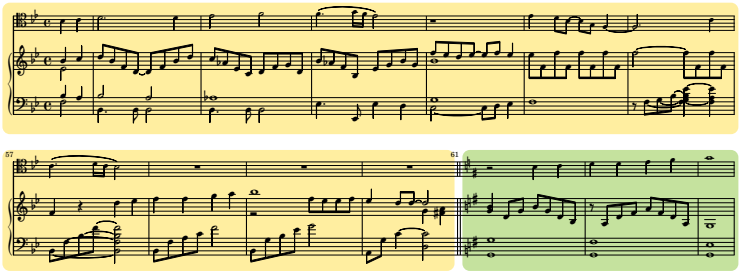
Understanding inner mechanisms in local key detection

modulation

John Rutter, *For the beauty of the earth.* *(Performance: Taipei Male Choir)*



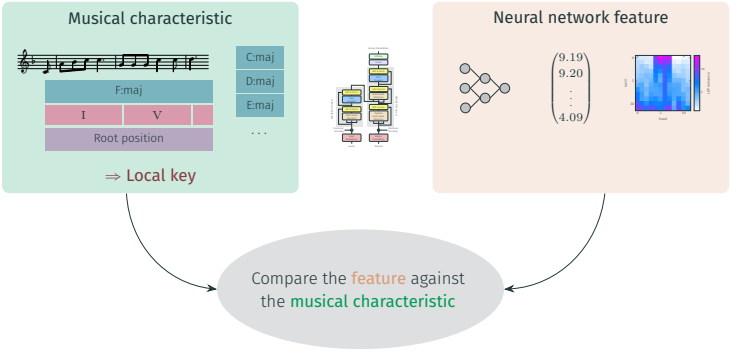
Understanding inner mechanisms in local key detection



Do the attention heads relevant for the annotation remain the same before and after the key change?

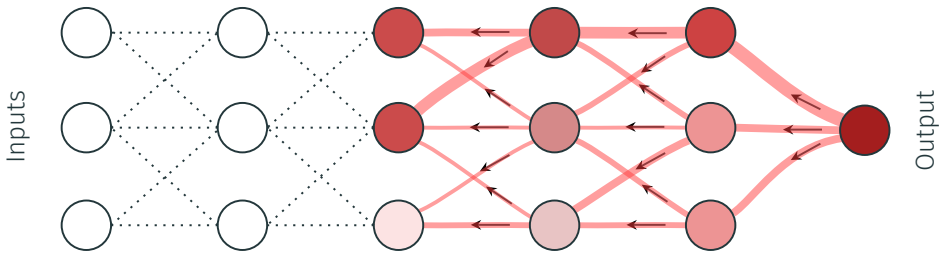
In the following, we will consider the model *only* trained on local key detection.

A methodology to interpret attention with regards to musical characteristics



Layer-wise relevance propagation (LRP)

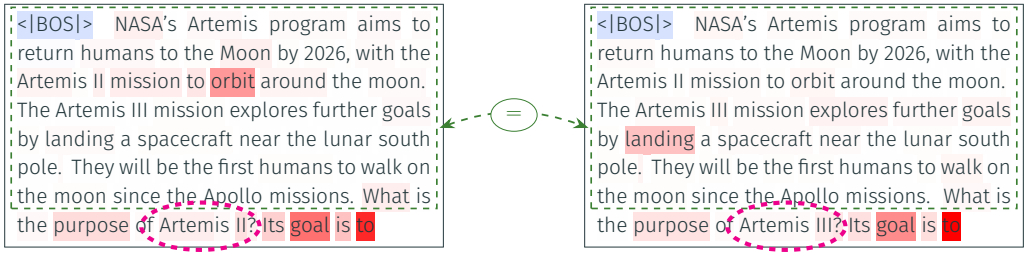
1. Consider a trained model on a 1-class classification task
2. Perform a forward pass and keep track of logits
3. LRP backward pass:



Montavon & al., *Layer-Wise Relevance Propagation: An Overview*, 2019

Layer-wise relevance propagation (LRP)

(In NLP) Model: Llama-3.2 (next word prediction)

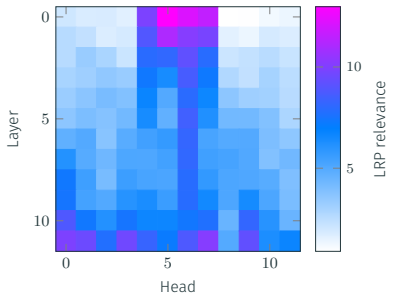


Achtibat & al., *AttnLRP: Attention-Aware Layer-Wise Relevance Propagation for Transformers*, 2024

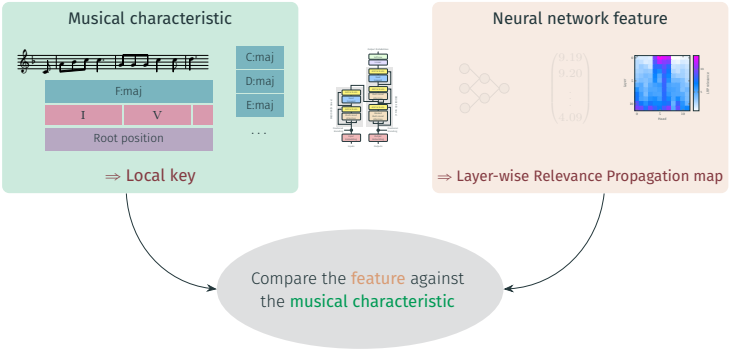
Layer-wise relevance propagation (LRP)

What if back-propagation is not performed up to the initial tokens?

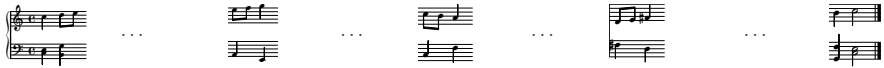
LRP map: contribution of each attention head to a prediction



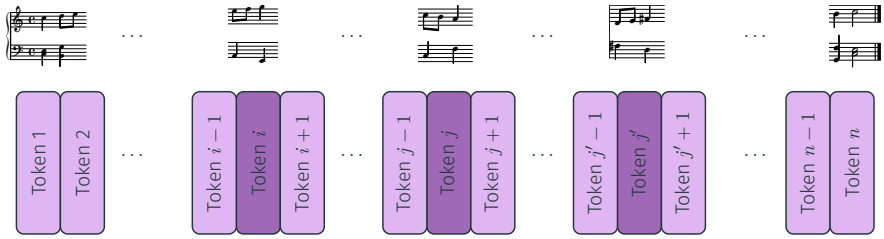
A methodology to interpret attention with regards to musical characteristics



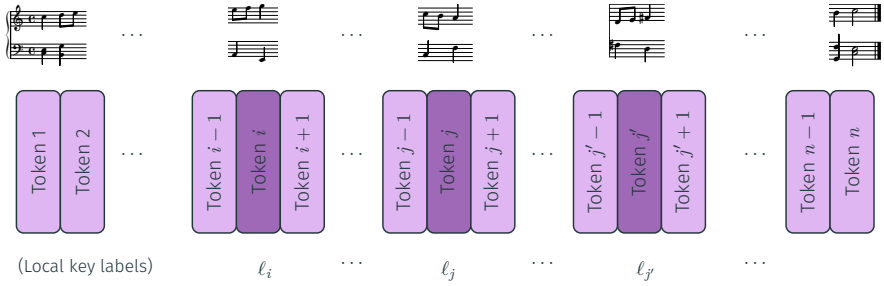
Attention head relevance for local key detection



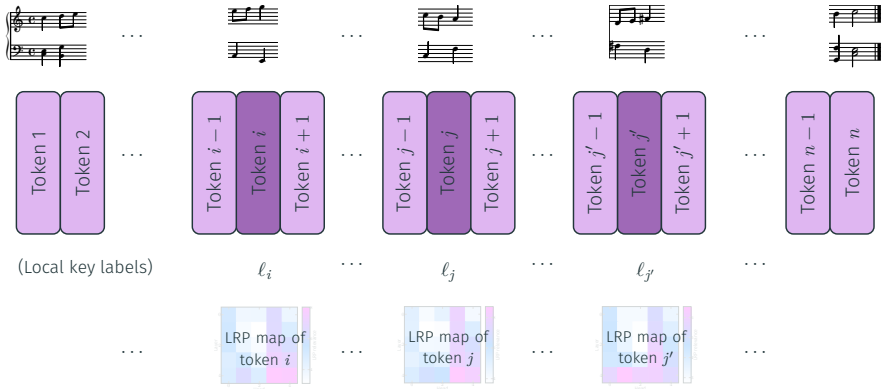
Attention head relevance for local key detection



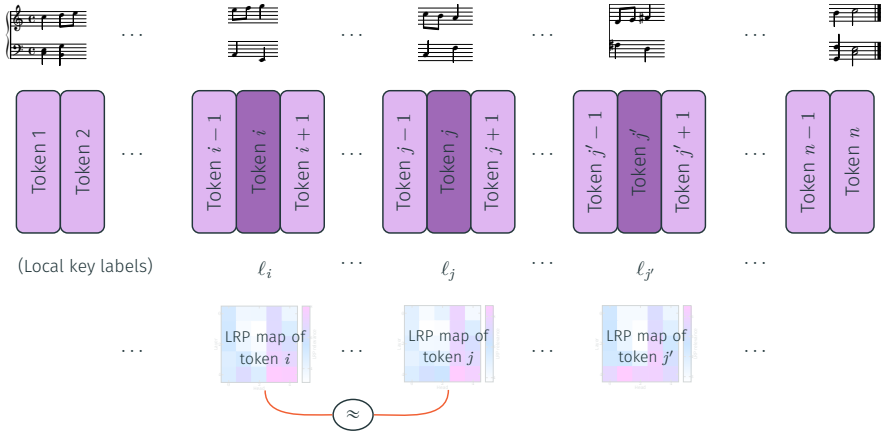
Attention head relevance for local key detection



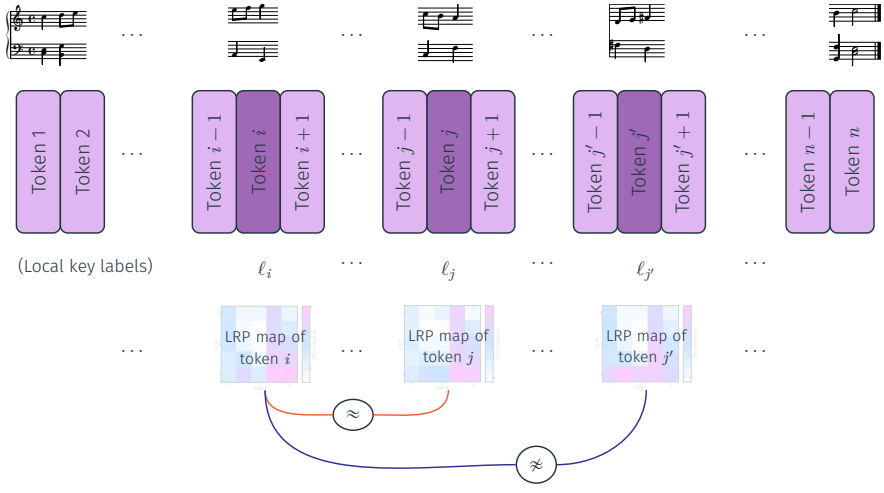
Attention head relevance for local key detection



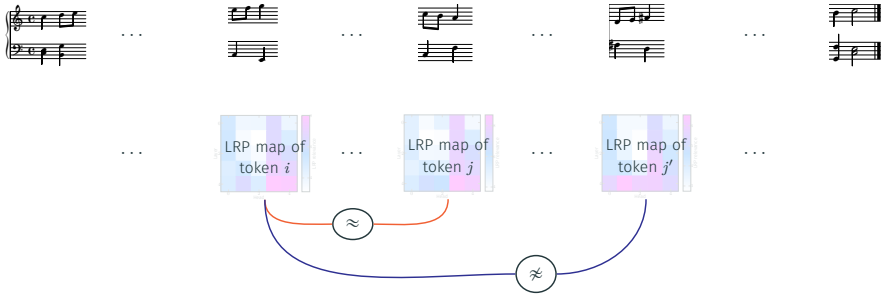
Attention head relevance for local key detection



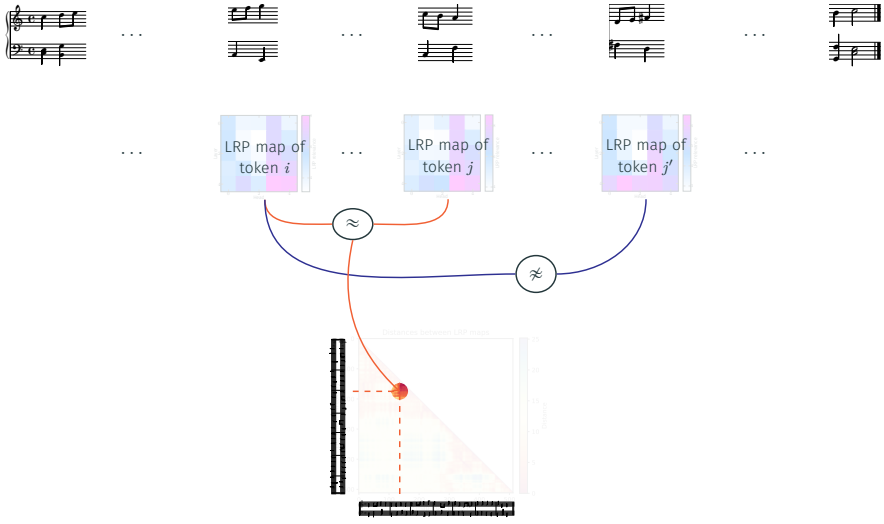
Attention head relevance for local key detection



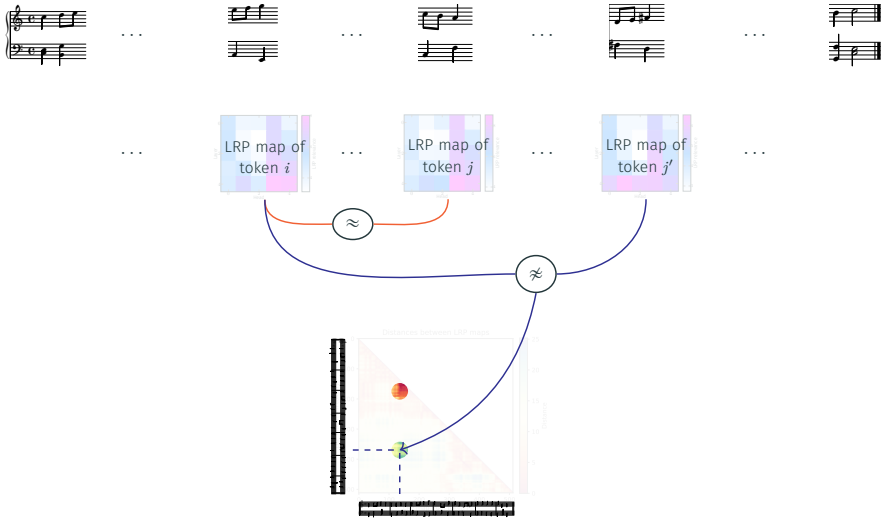
Attention head relevance for local key detection



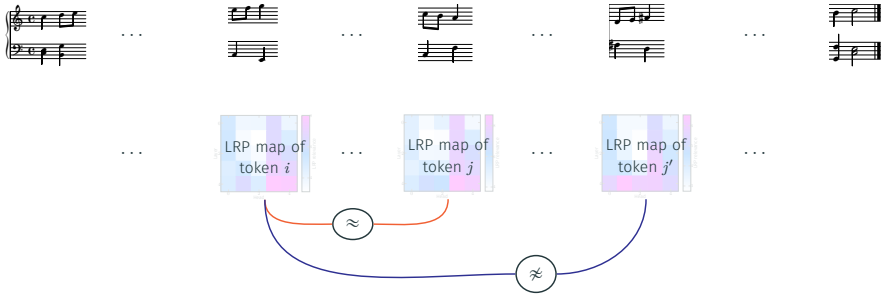
Attention head relevance for local key detection



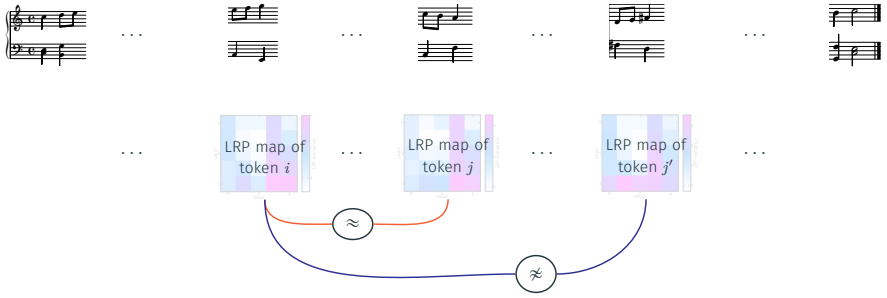
Attention head relevance for local key detection



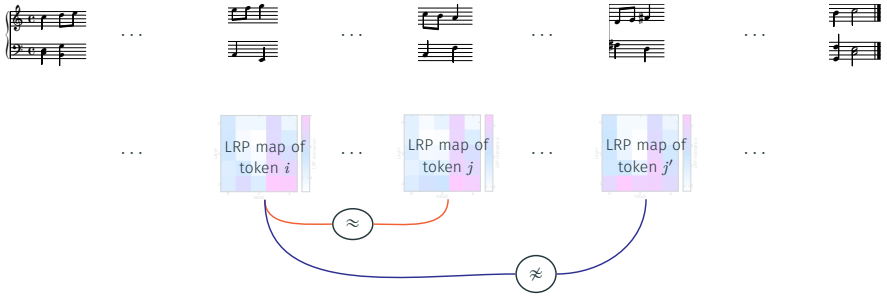
Attention head relevance for local key detection



Attention head relevance for local key detection

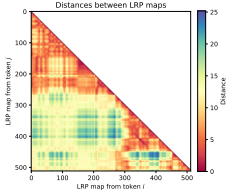


Attention head relevance for local key detection



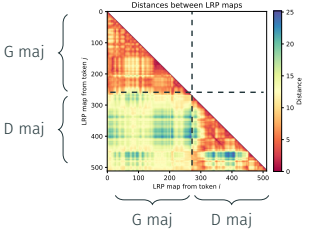
Attention head relevance for local key detection

Relating distances to labels



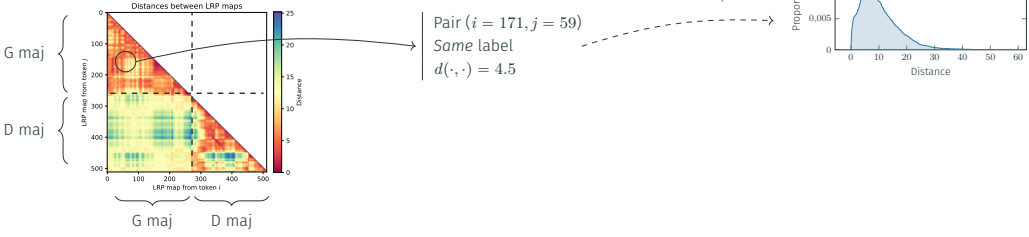
Attention head relevance for local key detection

Relating distances to labels



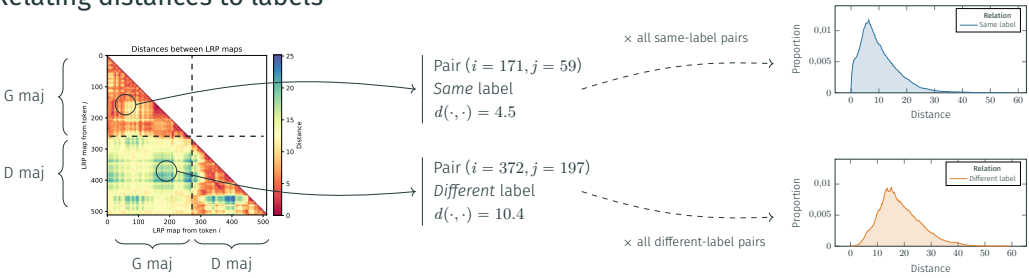
Attention head relevance for local key detection

Relating distances to labels



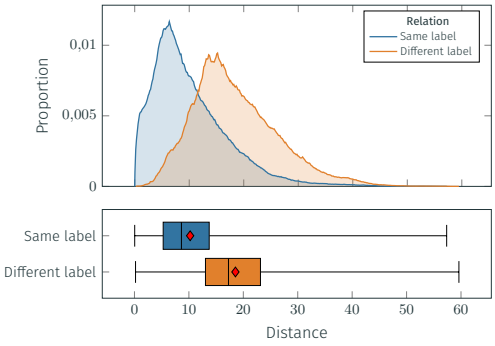
Attention head relevance for local key detection

Relating distances to labels



$\Rightarrow$ Comparing two distributions

Attention head relevance for local key detection



- The distributions between same vs. different labels are distinct (▲|▲)

⇒ **The relevant attention heads are different when changing keys.**

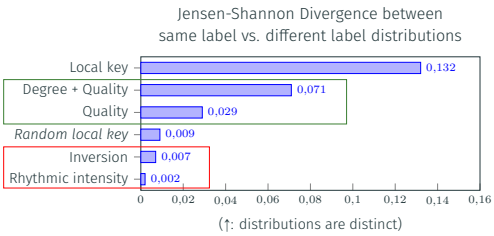
- Distances for same label pairs are lower than different label pairs (♦)

The attention heads relevant for the annotation of the *same local key* *are more similar* than for the annotation of different local keys.

Attention head relevance for local key detection

Comparing with other features

What if we consider other labels beyond local key?



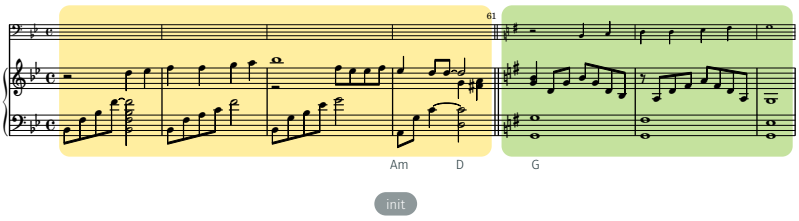
- Other harmonic features are also present in the model.
- Less relevant features for local key detection show less signal.

A model *only* trained on local key detection can capture *human-relevant* harmonic features it has not been trained on.

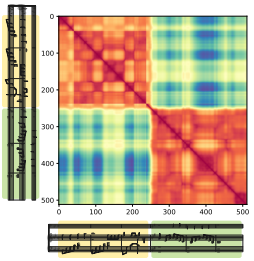
Attention head relevance for local key detection

What leads the model to “think” that a modulation has occurred?

Original modulation (B♭ maj → G maj)



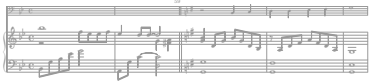
Distance between LRP maps

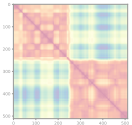


John Rutter, *For the beauty of the earth*.

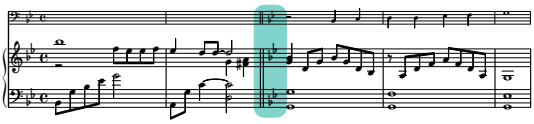
Attention head relevance for local key detection

Initial piece

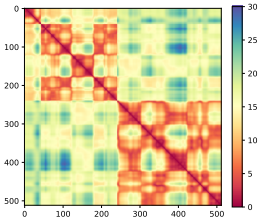




Keep the key signature



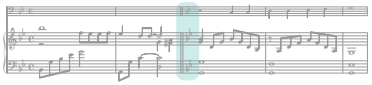
keep-key-signature

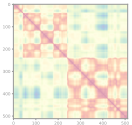


(⇒ *Modulation*
B♭ maj → G min)

Attention head relevance for local key detection

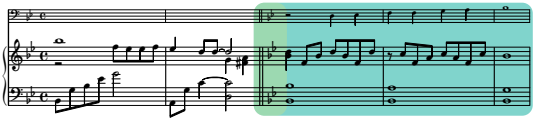
Last step:
Key signature kept



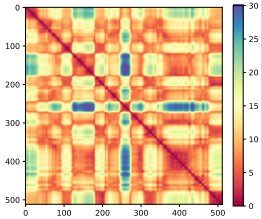


Transpose the modulated part to the initial key

(B♭ maj)

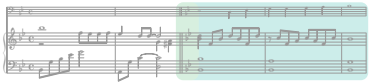


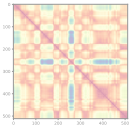
transpose



Attention head relevance for local key detection

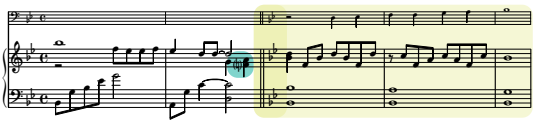
Last step:
Key signature kept
+ Transposition

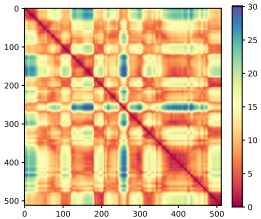




Remove the leading tone of the initial modulation

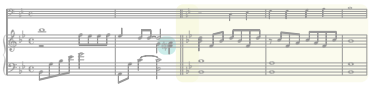
(F♯ in G maj)

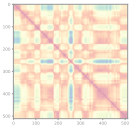




Attention head relevance for local key detection

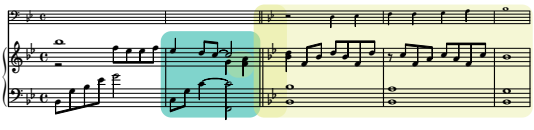
Last step:
Key signature kept
+ Transposition
+ No leading tone



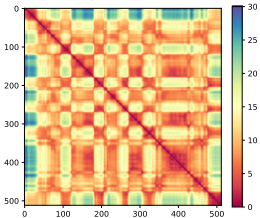


Re-write the initial modulation preparation

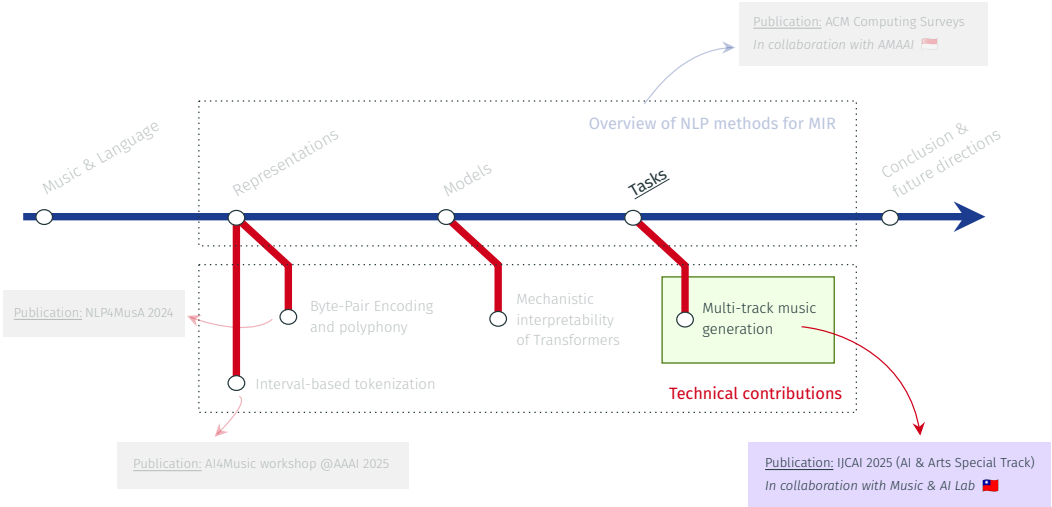
ii → V → I
G major (Am → D → G)
to B♭ major (Cm → F → B♭)



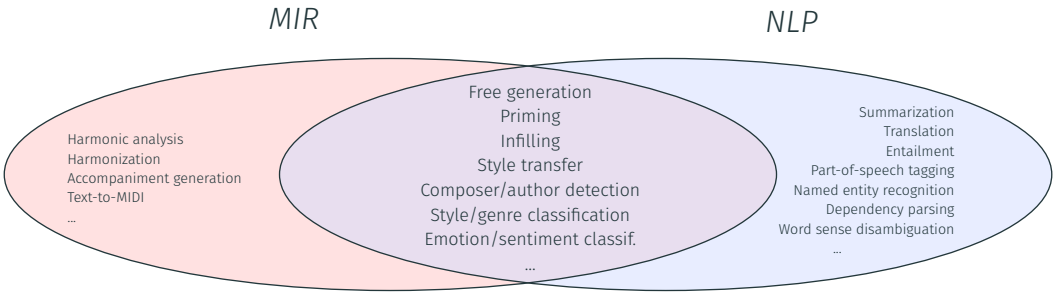
no-preparation



Outline – Tasks



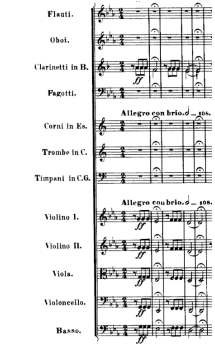
Tasks in NLP and MIR



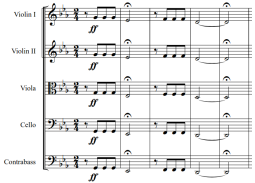
Re-orchestration: instrumentation, texture, melody

Instrumentation

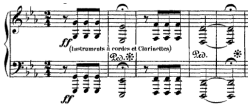
Symphonic orchestra



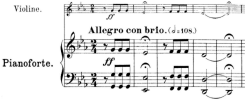
String orchestra



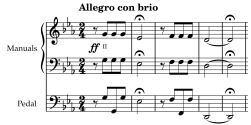
Piano solo



Piano - violin duet



Organ solo



L.v. Beethoven, *Symphony No. 5*, Mvnt. 1, Bars 1–5.

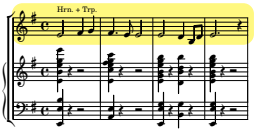
Orchestral texture

Melodic fidelity

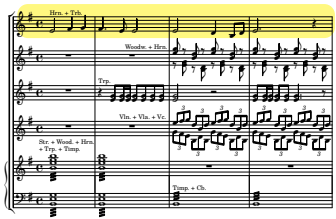
Re-orchestration: instrumentation, texture, melody

Instrumentation

Orchestral texture



dvorak9-10



dvorak9-279

A. Dvořák, *Symphony No. 9 "From the New World"*, Mvnt. 4.

Le & al., *A Corpus Describing Orchestral Texture in First Movements of Classical and Early-Romantic Symphonies*, DLfM 2022

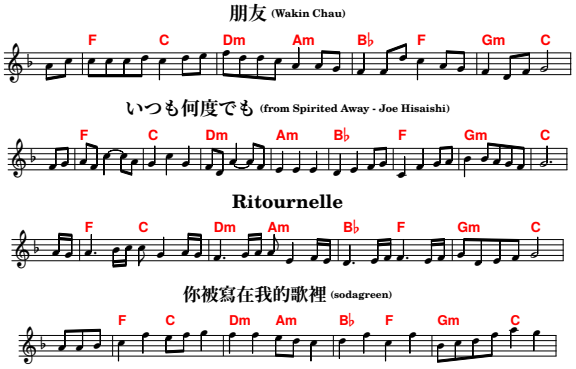
Melodic fidelity

Re-orchestration: instrumentation, texture, melody

Instrumentation

Orchestral texture

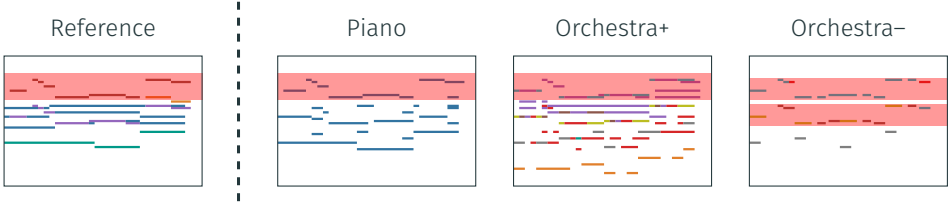
Melodic fidelity



("Pachelbel canon"-like chord progression: $\mathrm{I} - \mathrm{V} - \mathrm{vi} - \mathrm{iii} - \mathrm{IV} - \mathrm{I} - (\mathrm{ii}) - \mathrm{V}$)

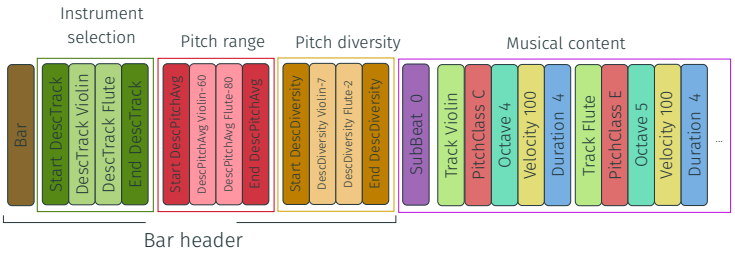
METEOR:

Melody-aware Texture-controllable Symbolic Music Re-orchestration via Transformer VAE



Le & Yang, *METEOR: Melody-aware Texture-controllable Symbolic Orchestral Music Generation via Transformer VAE*, IJCAI 2025 (AI & Arts Special Track). **In collaboration with Music & AI Lab (National Taiwan University).**

Tokenization & Textural controls



Bar-level

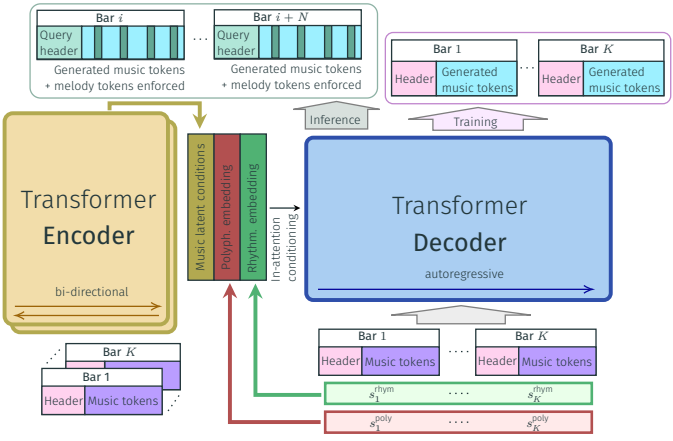
- Rhythmic intensity
- Polyphonicity

Bar+track-level

- Average pitch
- Pitch diversity

⇒ Derived statistically from the dataset

Model architecture



Wu & Yang, *MuseMorphose: Full-Song and Fine-Grained Piano Music Style Transfer with One Transformer VAE*, 2022

Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)
2. Choice of melodic instruments (e.g. `<Track-trumpet>`)
3. During inference

Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)

2. Choice of melodic instrumentsts (e.g. `<Track-trumpet>`)
3. During inference

Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)

2. Choice of melodic instructions (e.g. `<Track-trumpet>`)
3. During inference

Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)

2. Choice of melodic instrumentsts (e.g. `<Track-trumpet>`)
3. During inference

Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)

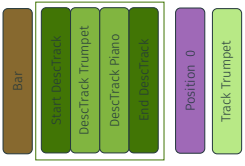
2. Choice of melodic instrumentsts (e.g. `<Track-trumpet>`)
3. During inference

Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)

2. Choice of melodic instrumentsts (e.g. `<Track-trumpet>`)
3. During inference

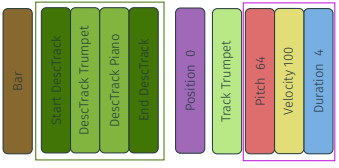


Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)

2. Choice of melodic instrumentsn (e.g. `<Track-trumpet>`)
3. During inference

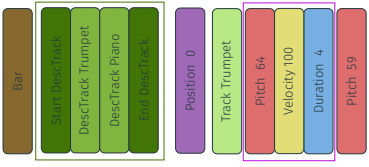


Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)

2. Choice of melodic instrumentsn (e.g. `<Track-trumpet>`)
3. During inference

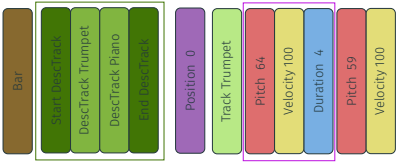


Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)

2. Choice of melodic instrumentsn (e.g. `<Track-trumpet>`)
3. During inference

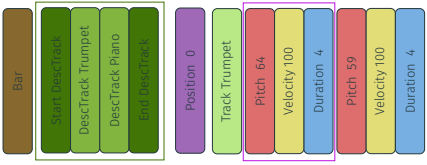


Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)

2. Choice of melodic instrumentsn (e.g. `<Track-trumpet>`)
3. During inference

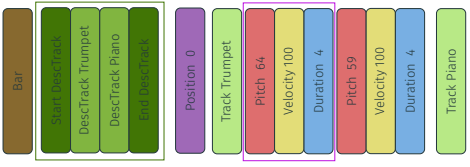


Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)

2. Choice of melodic instrumentes (e.g. `<Track-trumpet>`)
3. During inference

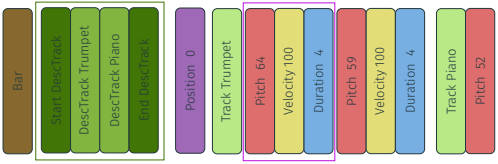


Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)

2. Choice of melodic instrumentals (e.g. `<Track-trumpet>`)
3. During inference

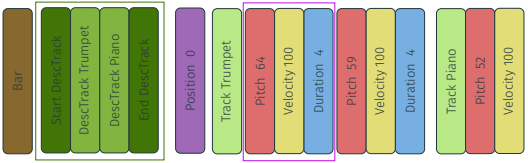


Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)

2. Choice of melodic instrumentss (e.g. `<Track-trumpet>`)
3. During inference

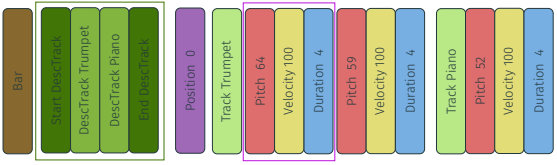


Melodic fidelity: inference guidance

Melody enforcing **during inference**.

1. Melody detection (bar-wise / instrument-wise skyline)

2. Choice of melodic instrumentss (e.g. `<Track-trumpet>`)
3. During inference

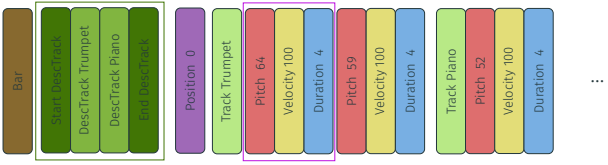


Melodic fidelity: inference guidance

Melody enforcing **during inference**.

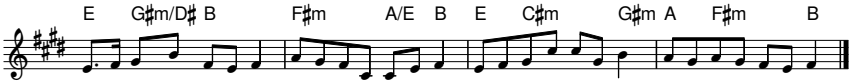
1. Melody detection (bar-wise / instrument-wise skyline)

2. Choice of melodic instrumentes (e.g. `<Track-trumpet>`)
3. During inference



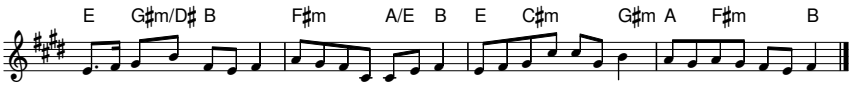
Lead sheet orchestration... for free?

Lead sheet: melody + chords

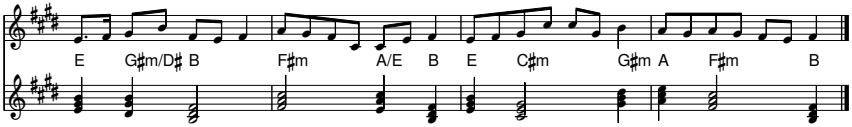


Lead sheet orchestration... for free?

Lead sheet: melody + chords



⇒ It is simply a low rhythmicity multi-track sample



⇒ **Lead sheet orchestration** = re-orchestration with higher rhythmicity (and higher polyphonicity) oboe-cello-harp-trio

Evaluation

Baseline models

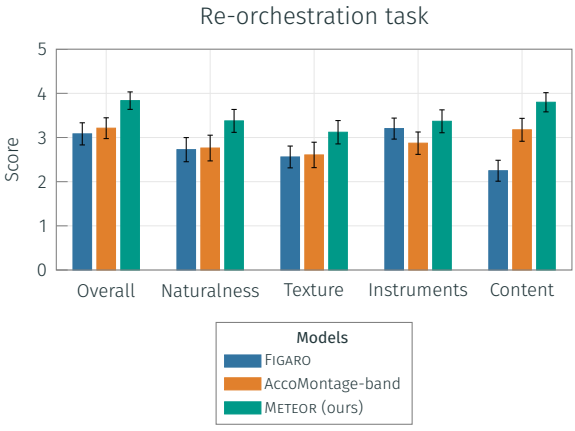
- FIGARO[2]: Multi-track music style transfer
- AccoMontage-band[3]: Lead sheet band arrangement

Objective (fidelity + controllability metrics) and **subjective** evaluation.

Criteria: Overall musicality ; Naturalness ; Texture fidelity ; Convincing use of instruments ; Content coherency.

[2] Rütte & al., *FIGARO: Generating Symbolic Music with Fine-Grained Artistic Control*, 2023
[3] Zhao & al., *Structured multi-track accompaniment arrangement via style prior modelling*, 2024

Subjective evaluation



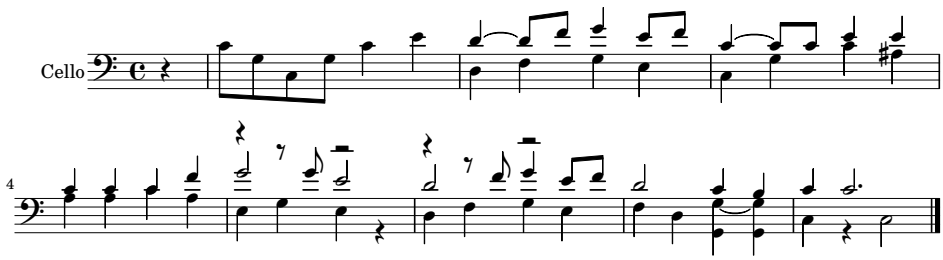
Piano solo → orchestra
Same texture
(Rachmaninoff - Prelude in G minor)

reference meteor

- **Content coherency**: FIGARO does not keep the melody figaro
- **Texture fidelity**: AccoMontage relies on a lead sheet transcription accomontage

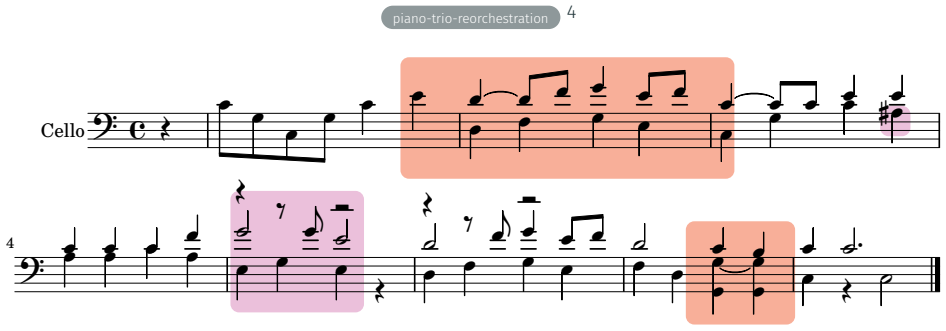
Towards symbolic music generation for *human use*?

piano-trio-reorchestration [4]



[4] **Reference**: lead sheet of Auld Lang Syne (*Ce n'est qu'un au revoir*).

Towards symbolic music generation for *human use*?



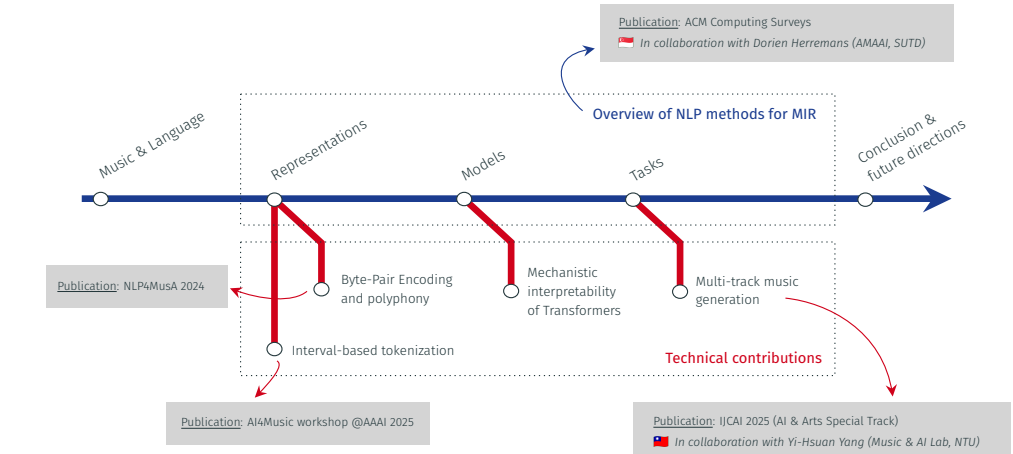
- Enharmonics, voice separation,...
- Awkward fingerings, articulations, breathings,...

→ How to ensure **humanly understandable** and **playable** generated music?

[4] **Reference**: lead sheet of Auld Lang Syne (*Ce n'est qu'un au revoir*).

Future directions & conclusion

Contributions, collaborations & publications



Other publications: Computational musicology (DLfM 2022) ; Music corpora (TISMIR 2025) ; AI for general public (Culture & Recherche 2024).

Future directions

Musical alphabet vs. text alphabet

- Is the Latin alphabet really the closest to musical alphabet?
- Can one still rely on MIDI to make *playable* music generation?

Data availability: text vs. symbolic music data

- Text data is released much more quickly than symbolic music
- Are larger models really the solution?

Future directions

Practices in the NLP field for MIR

- **Model explainability**: understanding the behavior of a model is essential, particularly in creative contexts.
  - *In NLP:* model-agnostic tools ; mechanistic interpretability
  - ISMIR 2025 Tutorial “Explainable AI for MIR”
- **Benchmarks**: how to compare models not developed for the same tasks?
  - *In NLP:* common metrics (BLEU for translation, ...) ; standardized benchmarks (GLUE, ...)
  - MIREX competition ; SMC Benchmark [Wang & al. ISMIR 2025]

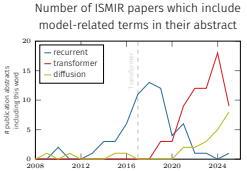
Towards further models for MIR

Future directions

Practices in the NLP field for MIR

Towards further models for MIR

- Transformers are still used, but have become “standard”
- Adapting models from other fields
  - *e.g.* diffusion models from image processing



⇝ What about a model tailored for music, beyond “simple” **adaptations** of other models?

Conclusion

How can one structure a MIR project?

- **Task**: how precisely is it musically defined?
- **Representation**: what motivates a particular choice?
- **Model**: why is it suitable for the chosen task and representation?

NLP as a *toolbox* for MIR

- NLP tools can be efficient to process symbolic music.
- Applying NLP tools for symbolic music should be first driven by *musical* questions.

Conclusion

How can one structure a MIR project?

- **Task**: how precisely is it musically defined?
- **Representation**: what motivates a particular choice?
- **Model**: why is it suitable for the chosen task and representation?

NLP as a *toolbox* for MIR

- NLP tools can be efficient to process symbolic music.
- Applying NLP tools for symbolic music should be first driven by *musical* questions.

Thanks for listening!

Modeling Symbolic Music with Natural Language Processing Approaches